

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 40.6668 Seconds
(without alignments)
2761.097 Million cell updates/sec

Title: US-09-647-544-2
Perfect score: 6106
Sequence: 1 MELPFTVHLPLPLVFLTGLC.....GFFAHKKIPDEKREKLEQ 1167

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:.*
1: Piri:.*
2: Piri:.*
3: Piri:.*
4: Piri:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1894.5	31.0	1180	2 A35854	integrin alpha-1 c
2	1887	30.9	1151	2 A45226	integrin alpha-1 c
3	1756	28.8	1181	2 A33998	integrin alpha-2 c
4	1754	28.7	1170	2 I45914	integrin alpha 2 s
5	1744	28.6	1178	2 S44142	VLA-2 protein homo
6	1105	18.1	1170	2 S03308	cell surface glyco
7	1097	18.0	1163	2 I56126	lymphocyte fuction
8	1085.5	17.8	1153	1 RWHU1B	cell surface glyco
9	1054	17.3	1163	1 RWHU1C	cell surface glyco
10	1022.5	16.7	1153	2 S00511	leukocyte surface
11	932.5	15.3	1179	2 A35312	integrin alpha-E c
12	738	12.1	1035	2 I58409	integrin alpha-9 c
13	732.5	12.0	1054	2 JC7294	alphap integrin -
14	690	11.3	1039	2 A41131	lymphocyte-Peyer's
15	684	11.2	1038	2 S06046	integrin alpha-4 c
16	661	10.8	1041	2 T31437	integrin alpha-4 c
17	655	10.7	272	2 A53348	integrin alpha-1 -
18	640.5	10.5	1137	2 JC5950	integrin alpha-7 c
19	602.5	9.9	1051	2 A40021	integrin alpha-3 c
20	600	9.8	1091	2 A41543	integrin alpha-6 c
21	598.5	9.8	1073	2 B36429	integrin alpha-6 c
22	596.5	9.8	1072	2 A38457	integrin alpha-6 c
23	594.5	9.7	1135	2 I61186	alpha-7 integrin -
24	584.5	9.6	1044	2 T10050	integrin alpha-v c
25	581.5	9.5	1053	2 I55534	VLA-3 alpha subuni
26	576.5	9.4	1044	2 I56516	integrin alpha-8 c
27	576.5	9.4	1051	2 A35761	cell surface glyco
28	574	9.4	1048	2 A27421	integrin alpha-5 c
29	573.5	9.4	1106	2 S38783	integrin alpha cha

30	570	9.3	1045	2	S60571	integrin alpha v c
31	566	9.3	1053	2	S44250	integrin alpha-5 c
32	563.5	9.2	1034	2	A36108	integrin alpha-v c
33	562	9.2	1049	2	A27079	fibronectin recept
34	543	8.9	1146	2	S40311	integrin - fruit f
35	540	8.8	1039	2	A34269	integrin alpha-2b
36	495	8.1	1037	2	A60163	glycoprotein I1b -
37	485	7.9	1226	2	S44824	F54F:1 protein -
38	475	7.8	1394	2	A29637	position-specific
39	445.5	7.3	1139	2	S28277	hypothetical prote
40	438	7.2	1086	2	T18523	integrin alpha cha
41	425.5	7.0	1115	2	T09433	integrin alpha cha
42	425.5	7.0	1115	2	T09403	integrin alpha cha
43	394.5	6.5	131	2	I47230	VLA-2 protein - pi
44	374	6.1	764	2	I36916	glycoprotein I1b -
45	303.5	5.0	604	2	I36917	glycoprotein I1b -

ALIGNMENTS

RESULT 1

A35854

integrin alpha-1 chain precursor - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 23-Oct-1990 #sequence_revision 13-Sep-1991 #text_change 09-Jul-2004

C/Accession: A35854; S11243

R/ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto,

J. Cell Biol. 111, 709-720, 1990

A/Title: Molecular cloning of the rat integrin alpha-1-subunit: a receptor for laminin a

A/Reference number: A35854; MUID:90338125; PMID:2380249

A/Accession: A35854

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1180 <IGN>

A/Cross-references: UNIPROT:P18614; GB:X52140; NID:G56493; PIDN:CAA36384.1; PID:G56494

C/Keywords: cell adhesion; cytoskeleton; transmembrane protein

F:170-345/Domain: von Willebrand factor type A repeat homology <VWA>

Query Match 31.0%; Score 1894.5; DB 2; Length 1180;
Best Local Similarity 36.5%; Pred. No. 1.6e-132;
Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;

Qy	13	LVLFTGLCSFNLDEHHPRLFPDPPEAFGYSVLQVGGQRMVLGVAPWDGSGDRGD	72
Db	19	LTVILGFCVSNVDVKNMSFSFVEDMFGYVQQYENEEGKWLIGSPLVGQPKARTGD	78
Qy	73	VYRCPVCGAHPNCAKGLGDYQLGNSSHPAV-----NMHLGMSLLETDCGGFMACAPL	127
Db	79	VYKCPVGRAMPVCVKLDLP-----VNTSIPNVTIENMTFG-STLVNPNNGGFLACGPL	133
Qy	128	WSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPTVMVIVLDGNSIYPWSEVQTF	187
Db	134	YAVRCGHLYHTTGICSDVSTFQVNSFAP-VQECSTQLDIVLDGNSIYPWSEVIAF	192
Qy	188	LRLVIGLFTDPEIQVGLVOYGESPVHMSLGDFTKEEVVRAAKNLSRREGRETAKQ	247
Db	193	LNLLKRMIDGPKQTVGIVQYGVNTHEFNLNKYSSTEEVLVAANKIGROGLQTTAL	252
Qy	248	AIWVACTEGFSQSHGGRPEARLLVVVTDGESHDEELPAALKACEAGRVTRYGIAVLGH	307
Db	253	GIDTARKEATFARGARRGVKVMVITDGESHONRYLKQVIOQCEBENQRFSAIALGH	312
Qy	308	YLRRQRDPSPFLRIRTIASDPDRFFNVNTDEAALTDIVDALDGRIFLGEGSHAENESS	367
Db	313	YNRCNLSTEFVEIKSIASEPTKEHFNVSDELATVIVKALGERIFALEATADQSAAS	372
Qy	368	FGLMSGIGFSTHRLKDGILFGVMGVADWGSVLWLGEGHRLFPFRMALEDEFPALQNH	427
Db	373	FEMMSQTGFSAHYSQDWVMLGAVGVDWNGTVVMQKQANQVIPHNTTFOTE--PAKQNE	430
Qy	428	--AAYLGYSVSSMLLRGGRRLFLSGAPFRFRHGKVIQAFOLKKDCAVRAVQSGEQIGSY	485

```
Db 431 PLASYLYTNSATIPGD-VLYIAGQPRYNHTGQVITYKM-EDGNINILQTLGEQIGSY 488
Qy 486 FGSELCLDTRDQTTDVLVAAFMFLGPONKGTGRVYVYLVGQOQLTLTQGLTLP---- 541
Db 489 FGSVLTTIDDKSYTDLVLVGAAPMYGTEKEGQKYVYAV-NQTFEYQMSLEPIRQT 547
Qy 542 -----EPPQDARFGFAMGALPDLNQGDFADVAVGAPLEDGHQGALYLYHG 586
Db 548 CCSLKDNSCKTENKNEPCGARGFATAA VKDLNVDFNDVWIGAPLEDHAGAVIYHG 607
Qy 587 TQSGVRPHPAQRIAAAMPHALSYPGRSVGRDLDDGDDLVDAVGAQGAAILSSRPV 646
Db 608 SGTIREAYAQRIPSGGDKTLKFFQOSIHGEMDLNGDGLTDTVITIGLGAALFWARDA 667
Qy 647 HLTPSLEVTQAISVVDRCRRQGEAVCLTAALCFQVTSKTPGRMBHQFMFRTASLDE 706
Db 668 VVKVTMNFENKVNQKNCRVEGETVCINATCFHVKLKSKEDSIYEADLQVRYTLD 727
Qy 707 WTAGARAAPDGGQRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRLPVALTFTFALDNTTK 766
Db 728 LRQISRSFFSGTQERKIQ--NITVRESECIHRSFYMLDKHDFQDSVRVTLDF---NLTD 782
Qy 767 P--GPVLNBSGPTSQIKLFPSCDGPDCPDNECVTDLVQVMDIRGSKAPFVVRGGRKV 824
Db 783 PENGVPVLDALPNSVHEHIFPAKDCGNKRCISDLTLNVST---TEKSLLIIVKSQHDKF 838
Qy 825 LVSTTLNKRKENAYNTSLSIIFSRNLHLASLTP-QRESPIKVECAAPSAHARLCSVGHV 883
Db 839 NVSLTVKXKGDASNTVTVQHSNLLIFSGIBELQKOS-----CESNQNTICRVGYPP 891
Qy 884 FQTGAKYVTLLEPFSCSSLLSQVFGKLTASSDSLRNGTLQENTTAQTSAYIQYEPHLF 943
Db 892 LRAGETVTKIIPQFNTSHLSENAI IHLSATSDSEEPLESINDNEVNISSIPVKEVGLQF 951
Qy 944 SSESILHYEHPVGTLP-----VGPQPEKTLRVQNLGCVVSGLIISALLP-AVAH 996
Db 952 YSSASEHHISVAANETIPEFINSTEDIGNEINVFTIRKGRHPPMPELQLSISPPNLTAD 1011
Qy 997 GGNVFLSLQSIVITNASCIVQNLTEPPG-----PPVHPPELQHTNRLNGSNTQCQVVR 1050
Db 1012 GYPVLYPIGWSSSDNVNCRFASLBDPFGINSKQKQTKSKSEVLKRGHIDCSSTCGVATI 1071
Qy 1051 HLGQAKG-TEVSVGLRLVHNEFFRRAKFKSLTVVSTFELGTBEGSVLQJTEASRWSES 1109
Db 1072 TCSLLPSDLQSVNVL--LLWKPTFIRAHFSSLNLTLRGELKSENSS-LTLSSNRKREL 1128
Qy 1110 LLEVQOT-RPILISLWILIGSVLGGLLALLVFLWKLWGLFFPAHKKIPBEKEE 1163
Db 1129 AIQISKDGLPGRVPLWILLSAFAGLLLLILLIALLWKIGFF---KRPLKKOMEK 1180

RESULT 2
A45226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45226
R:Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Cross-references: UNIPROT:P56199
A:Experimental source: hepatoblastoma cell line HepG2
A>Note: sequence extracted from NCBI backbone (NCBIP:124326)
F:142-317/Domain: von Willebrand factor type A repeat homology <VWAI>

Query Match 30.9%; Score 1887; DB 2; Length 1151;
Best Local Similarity 36.3%; Pred. No. 5.4e-132;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;
```

```
Qy 23 FNLDHHPRLFPGPPEAEFGYSVLQHVGGQRMMLVGAPWDGPGSRRRGDGVYRCPVGAH 82
Db 1 FNVYKNSMTFSGVEDMFGYTVQYENBEGKWLIGSPVQPKNRTGDVYKCPVGRGE 60
Qy 83 NAPCAKGHLDYOLGNSHPAV-----NMHGLMSLLETGDDGDFMACAPLWSRACSSVF 137
Db 61 SLPCVKLDLP----VNTSIPNVTENKMTFG-STLVNPNNGGFLACGPLYAYRCCGHLHY 115
Qy 138 SSGICARVDASFPQOGLAPTAQRCPYMDVIVLDGNSIYPMWSEVQTFRLRLVGKLEI 197
Db 116 TTGICSDVSPTFQVNNSIAP-VQSCSTQDQIVIVLDGNSIYPMWDSVTAPLNDLLKRMDI 174
Qy 198 DPEIQIVGLVQYGESVPHWESLGDFTKEBVRAAKNLNRREGRETQTAQINVACTEGF 257
Db 175 GPQKTQVGIQYGENVTHEFNLNKYSSTEVLVAKKIVQRCGRQWTALTGTARKEAF 234
Qy 258 SQSHGRRPBAARLLVVVTGESHGDBELPAALKACBAGRVTYRGIYAVLGHYLRQRDPSS 317
Db 235 TEARGARRGVKVMVITDGHSHNHLKVKIQDCEDENIQRFSAIILGSYNRLSTEK 294
Qy 318 FLREITIASDPPERFENVTDEAALTDIVDALGDRIEGLGSHAENESSFGLEMSQIGF 377
Db 295 FVEEIKSIASEPTEKHFENVSDELAVTIVKTIGERIFALEATADQSAASFEEMSQTGF 354
Qy 378 STHRLKDGILFGMVGAYDMGGSVLMLEGGRHLPFPRMALEDEPPALQNHAAVLYGSVSS 437
Db 355 SAHYSQDMVLMGAVGAYDMNGVTVMQKASQIIPRNTTFNVESTKNEPLASLYGTVNS 414
Qy 438 MLRGGRRPLSGAPRHRGKVIAPQLKKGAVRVAQSLQGGQISYFGSELCPLDTR 497
Db 415 ATASSGDVLYIAQPRYNHTGVIIYRM-EDGNIKILQTLGSGQISYFGSILTTTDDK 473
Qy 498 DGTDDLVAAPMFLGPONKGTGRVYVYLVGQOQLTLTQGLTLP----- 541
Db 474 DSNTDILLVGAPMYMTEKEBQKVYVYAL-NQTFEYQMSLEPIKQTCSSQHNSCTT 532
Qy 542 ---EPQDARFGFAMGALPDLNQGDFADVAVGAPLEDGHQALYLYHGTSQVVRHPAQR 598
Db 533 ENKNEPCGARFGTAIAAVKDLNLDGNDIVIGAPLEDHGGAVIYHSGKTKIRKVAQR 592
Qy 599 IAAASMPHALSYGRSVGDRDLDDGLVDVAVGAQGAAILSSRPVHLTPSLEVTPOA 658
Db 593 IPSGDGLKTLKFFQSIHGBMDLNGDGLTDTVITIGLGAALPMSRDVAVVVKVMTPEPNK 652
Qy 659 ISVVQRDCRRRQGEAVCLTAALCFQVTSRTPGRMDHQFYMRFTASDEWTAGARAFDGS 718
Db 653 VNTQKNCHMEGKETVCINATVCFEVKLKSKEDTIYEADLQYRVTLDSLRQISRSFSGT 712
Qy 719 GQRLSPRRRLKLSVGNVTCQQLHFHVLDTSDYLRPVALTFTFALDNTTKP--GPVLNBSG 776
Db 713 QERKVQR--NITVRKSECTKHGFYMLDKHDPQDSVRITLDF---NLTDPENGFVLDLSP 767
Qy 777 TSOKLVPSKCGCPDNECVTDLVQVMDIRGSKAPFVVRGGRKVLVSTTLNKRKEN 836
Db 768 NSVHEVTPFAKDCGNKEKCIISDLSHV----ATTEKDLIVRSQNDKFNVSILTAVNTKDS 823
Qy 837 AYNTSLSIIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPFQGAQVTLLEF 896
Db 824 AYNTRIYHVSPLNVFSGI----EATQKOSC--ESNHNITCKYGVPLRRGEMVTKILF 877
Qy 897 EFGCSLLSQVFGKLTASSDSLRNGTLQENTTAQTSAYIQYEPHLFSSSESTLHYEVHP 956
Db 878 QFNTSYLMENVTIYLSATSDSEPEPTELSNVNVISSIPVKEVGLQFYSSASEVHISIAA 937
Qy 957 YGTLP-----VGPGEFETKTLRVQNLGCVVSGLIISALLPAVAGGN---YFLSLSQV 1007
Db 938 NETVPEVINSTEDIGNEINIFYLIRKSGSPMPELKLISFPNMTSGYPLVPTGLSS- 996
Qy 1008 ITWNASCIQNLTEP-----PGPPVHPPELQHTNRLNGSNTQCQVVRCHLQGLAKGTEV 1061
Db 997 -SENANCRPHIFDPDFPSINSKKQMTSTDLHKKRGTLDCNCTCKPATITCMLTS-SDISQV 1054
```

QY 1062 SVGLRLVHNEFFRRRAKFKSLTVVSTELCTEGSVLQLTASRWSLSLEVQVTP-RPIL 1120
 Db 1055 NVSL--LLWPTFKTSYFSSNLTIRGELSENAS-LVLSSNQKRELAIQISKDGLPGR 1111
 QY 1121 ISWLILGCVLGLLLALLVFCILWKLGFFAHKIPIPEEKREE 1163
 Db 1112 VPLWVILLSAPAGILLMLLILALWKIGFF---KRPLKKKMEK 1151

RESULT 3
 A33998
 N;Integrin alpha-2 chain precursor - human
 N;Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 chain; Species: Homo sapiens (man)
 C;Date: 30-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 09-Jul-2004
 C;Accession: A33998; B56793; A53117
 R;Takada, Y.; Hemler, M.E.
 J. Cell Biol. 109, 397-407, 1989
 A;Title: The primary structure of the VLA-2/collagen receptor alpha (2)-subunit (platelet)
 A;Reference number: A33998; PMID:89308879; PMID:2545729
 A;Accession: A33998
 A;Molecule type: mRNA
 A;Residues: 1-1181 <TAK>
 A;Cross-references: UNIPROT:P17301; GB:X17033; NID:q33906; PIDN:CAA34894.1; PID:q33907
 A;Note: the authors translated the codon GAT for residue 802 as Gln, GTC for residue 803
 R;Cacimel, B.; Parentier, S.; Leung, L.L.; McGregor, J.L.
 Biochem. J. 279, 419-425, 1991
 A;Title: Separation of important new platelet glycoproteins (GPIa, GPIC, GPIc*, GPIIa and GPIb)
 A;Reference number: A56793; PMID:92061944; PMID:1953640
 A;Accession: B56793
 A;Molecule type: protein
 A;Residues: 30-43 <CAR>
 A;Experimental source: platelet
 R;Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Teung, Y.L.; Gafford, A.
 J. Biol. Chem. 269, 463-469, 1994
 A;Title: The human alpha-2 integrin gene promoter. Identification of positive and negative regulatory elements
 A;Reference number: A53117; PMID:94103255; PMID:8276836
 A;Accession: A53117
 A;Molecule type: DNA
 A;Residues: 1-16, 'V', 18-21 <ZUT>
 A;Cross-references: GB:L24121; NID:g400342; PIDN:AAA16619.2; PID:g4583535
 A;Note: authors translated the codon GTA for residue 17 as Leu
 C;Genetics:
 A;Gene: GDB:ITGA2; CD49B
 A;Cross-references: GDB:128031; OMIM:192974
 A;Map position: 5q11.1-5q11.2
 C;Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein
 F;1-29/Domain: signal sequence #status predicted <SIG>
 F;30-1133/Domain: extracellular #status predicted <EXT>
 F;172-347/Domain: von Willebrand factor type A repeat homology <VWA2>
 F;1134-1154/Domain: transmembrane #status predicted <TM>
 F;1155-1181/Domain: intracellular #status predicted <CYT>
 F;105,112,343,432,460,475,699,1057,1074,1081/Binding site: carbohydrate (Asn) (covalent)

Query Match 28.8%; Score 1756; DB 2; Length 1181;
 Best Local Similarity 34.3%; Pred. No. 3.4e-122;
 Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

QY 11 LPLVPLTGL-----CSPFNLDHHPRLPFGPPEAFEGYSVLQHVGGQGRWMLVGAPWD 63
 Db 11 LPLLLVLLSQILNCLLAYNVGVLPEAKIFGSPSEQFGYAVQVFINPKGNWLLVGSFWS 70
 QY 64 GPSGRRGDVYRCPVGGAAHNAFCARGLH-GDYQLGNSHPAVNMHLGMSLLTETDGDGGM 122
 Db 71 GFPENRMDVYKCPV-DLSTATCEKLNLTSTSPNVTEMTKNTNLSGLILTRNMGTFGL 129
 QY 123 ACAPLWSRACSSVPSSGICARVDASFQPSLAPTAQRCPTYMDVIVLDCGSSITPWS 182
 Db 130 TCGPLWAAQCCNQYVITGVCSDISPDFQLSASFSPATQCFSLIDVWVVCDESNIYPWD 189
 QY 183 EVQTFRLRLVGLKFLTDPQIQVGLVQYGESPVHEWSLGDFTKERVVRAAKNLSRRGRE 242
 Db 190 AVKNFLEKFGVGLDIGNTKTQVGLIQYANNPRVFNLTNTYKTEEMIVATSTQSYGGDL 249

RESULT 4

I45914

Integrin alpha 2 subunit - bovine (fragment)

QY 243 TKTAQAIMVACTGFSQSHGCRPEAARLLVVTVDGESHGDEELPAALKACAGRVTRYGI 302
 Db 250 TNTFGAIQYARKYAYSAAAGRRSATKVMVWVVDGESHGSMKLVAVDQCNDHNDILRFGI 309
 QY 303 AVLGHYLRRQRDPSSFLUREIRTIASDPDEFNFNVNTEAALTIDIVDALGDRIFGLESHA 362
 Db 310 AVLGYNLRNALTWNLIKKEIKATASIPTEYFFNVSDAALLEKAGTGLGQIFSIETG-V 368
 QY 363 ENESFGELEMSQIGFST--HRLKDGILFGMVGAYDMGGSVLW-LEGGRHLRPPPRMALEDE 419
 Db 369 QGDNFMQMSQVGSFADSYSSQNDILMLGAVGAFGWSGTIVQKTSHGHLIFP-----KQA 423
 QY 420 FPPALQ--NHAAYILGYSVSMILRGRRFLSGAPRFRHGKVIAPOLKDGAVRVAQSL 477
 Db 424 FDQILODRNHSSLYGVSAVA-ISTGESTHFVAGAPRANYTGOIVLYSVNENGNIYQAH 482
 QY 478 QGQIGSYFGSELCLPLDTRDGTDLVLLVAPMFLGPONKETGRVYVY-----LVGQOQL 532
 Db 483 RGOIGSYFGSVLCSDVDNDKTTDVLVGAPMTMSDLKKBEGRVYLTFTIKKGLGQHQF 542
 QY 533 LTLQGLQPEPPQDAREFGFANGALPDLNQDGFADVAVGAPLEDGHQGLALYHGTQSGVR 592
 Db 543 --LEG---PSGIENTRFGSAIALSDINMGDFNDVVGSPLENQNSGAVIYNGHQGTIR 597
 QY 593 PHPAQRIAAA--SMPHALSIFGRSDGRDLDDGDDLDVAVAGQAAILSSRPVHLTP 650
 Db 598 TKYSQKILGSDGAFRSHLQYFGRLDGYGLNGDSITDVSIGAFGVQVQLMSQSIADVAI 657
 QY 651 SLEVTQPAISVQORDCHRRRQEAVALCFAQVTSRTPGRDWHQFVMTFTASIDEWTAG 710
 Db 658 EASFTPEKITLVNKNQ-----ILKLCFSAKFR-PTKQNNQVAIVNITLDA----- 704
 QY 711 ARAAFDGSQORLSPRL-----RLSVGNV-----TCEQLHFHVLDTSDYLRVALTVT 758
 Db 705 -----DGFSSRVTSRGLFKENNERCLQKNVNVQAQSCPEHIIYIQPSDVVNSDLRD 759
 QY 759 FALDNTTKPG--PVLNEGSPISOKLVPSKDGPDNECVTDILVQVNMIDIRGSRKAPV 816
 Db 760 ISLEN-----PGTSPALEAYSETAKVPSIPFHKDCGEDGLCISDLVDVR-QIPAAQEQPFI 815
 QY 817 VRGRRKVLVSTLTENKENAYNTLSIIFSRNHLASLTPQRESPI---KVEC-AAPSA 872
 Db 816 VSNQNKLTFTVLNKNRESAYNTGIVDFSENLFPAFS-----LPVDGTEVTCQVAAQ 871
 QY 873 HARLCSVGHVFOFGKAVTFLLEFEFSCSLLSQVFGKLTASSDSLERNGTLOENTAQTS 932
 Db 872 KSVACDVGYPALKREQQVTTINFDNLQNLQNASLSFOALSSESQENKA--DNLVNLK 929
 QY 933 AYIQYEPHLLFSSSESTLHRYEVHPYGLP-----VGPGEFETTLRVQNLGCVVSGLI 986
 Db 930 IPLLVDAAEIIHLTRSTNINFEISSDGNVPSIVHSFEDVGPKFIFSLKV-TTGSVPVSNAT 988
 QY 987 ISALLPAVAGNVPFLSLSOVITNNASCIQNLTEPP-----GPPVHPEELOHTNRL 1038
 Db 989 VIHIPOYTKENKPLMYLTGTQTDKAGDISCNADINPLKIQOTSSSVSFKSENFRHTKEL 1048
 QY 1039 NGSNTOCVVVRCHLQGLAKGTEVSGLLRLVHNEFFRRAKFKSLTVVSTFELGTEGSVL 1098
 Db 1049 NCRTASCSNVTCLWKDVHMKGEYFVNVVTRIWNGTFASSITFQVLTAAAEINYNPEIY 1108
 QY 1099 QLTEASRWSLSLEVQVTRPILISLWI-----LIGSVLGLLLLLALLVFLCLW 1145
 Db 1109 -----VIEDNTVTIPLIMKPKDEKAEVPTGVIIGSIIAGIILLALLVAILW 1154
 QY 1146 KLGFPAHK-----KIPES 1158
 Db 1155 KLGFKKKYEKMTKNPDE 1172

C;Species: Bos primigenius taurus (cattle)
 C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
 C;Accession: I45914
 R;Kamata, T.; Puzon, W.; Takada, Y.
 J. Biol. Chem. 269, 9659-9663, 1994
 A;Title: Identification of putative ligand binding sites within the I-domain of integrin
 A;Reference number: A54402; MUID:94193647; PMID:7511592
 A;Accession: I45914
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1170 <KAM>
 A;Cross-References: UNIPROT:P53710; GB:L25886; NID:9439695; PIDN:AB59255.1; PID:9439696
 F;161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 28.7%; Score 1754; DB 2; Length 1170;
 Best Local Similarity 33.9%; Pred. No. 4.7e-122; Indels 102; Gaps 30;
 Matches 411; Conservative 235; Mismatches 465; Indels 102; Gaps 30;

Qy 9 LFLPLVLTGL---CSPFNLDHHPRLFPQPPAEFGYSVLQHVGGQRMWLVGAPWDGP 65
 Db 2 LQLVLVPSQILNCVAYNVGLPKAKIFSGPSSBQFGYAVQFINPKGNMLLVGSPWSGF 61
 Qy 66 SGDRRGDYYRCPVGGAHNAPCAKGLH-GDYQLGNSSHPAVNMHGLMSLLETDDGGFPMAC 124
 Db 62 PKRMGDVYKCPV-DLSTTTCCKLNLTSTSMGNVTETKMTNMSLGLTLTRNVGTGGFLTC 120
 Qy 125 APLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMYDVIIVLDGNSIYPMSEV 184
 Db 121 GPLWAQCGSQYTTGVCSDVSPDFQRTSFAPAVQTCPSFIDVVVVCDSNSIYPMWAV 180
 Qy 185 QTLRLRLVGLKFLIDPEQIVGLVQYGESPVHESLGDFTKBEVYVRAAKNLSRREGRETK 244
 Db 181 KNFLKPVQGLDIPGTTKQGLIYANNPRVFNLTFSKDEMIKATSTQFYGGDLTN 240
 Qy 245 TAQAIWVACTEGFSQSGHGRPEARLLVVVTDGESHGDBELPAALKACAGRVTRYGIIV 304
 Db 241 TFKAIQIARDTAYSTAAGRPGATKVMVVVTDGESHGSKLKAVIDQCNKDNLIRFGIIV 300
 Qy 305 LGHYLRQRDPSPFLRIRTIASDPDRFFNFVTDAAALTDIVDALGDRIPLGEGSHAEN 364
 Db 301 LGVYLRNALDTQNLKIKAIASIPTEHPFNVSDEADLLEKAGTIGEQIFSTEGT-VQG 359
 Qy 365 ESFGLEMSQIGFSTHRLKDG--ILFGWVGAYDWGGSVLW-LEGHRLFPFRMALEDFP 421
 Db 360 GDNFQEMSGVGSAAEYSPQNNILMLGAVGAYDWSGTVMQKTPHGLHIFS-----KQAFE 414
 Qy 422 PALQ--NHAAYLIGYSVSSMLLRGRRLFLSGAPRFRHGRKVIAPQLKKDGAIVRAQSLQ 479
 Db 415 QILQDRNHSSYLGYSVAS--ISTGNSVHFVAGAPRANTYTGQIVLYSVNENGNTVVISQ 473
 Qy 480 EQIGSYFGSLCPDLDTRDGTDLVLAAPMFLGPQNKETGRVYVYLVGQSLTLTGTL 539
 Db 474 DQIGSYFGSVLCAVDVNDKITDVLVGA PMYMDLKEBGRVLYFTITKILNWHQFLE 533
 Qy 540 QPBPPODARFGFAMGLPDNLNQGFADVAVGAPLEDHGHQALYLYHGTQSGVRPHPAQRI 599
 Db 534 GPNGLENARFGSAI AALSDINMGDFNDVI VGSPLENQNSGAVYIYNGHEGMIRLRYSQI 593
 Qy 600 AASAMPHA--LSVYGRSVGRDLDDGDDVDVAVGAQGAAILLSRPIVHLTSLVTPQ 657
 Db 594 LGSRAFSHLLQVYGRSLDGGDLNGDSITDVSAGAFQVQVQLMSQSIADVSDASTPK 653
 Qy 658 AISVQRDCRRRQGEAVCLTAALCFQVTSRTPGRDHDFYMRFTASLDWETAG----ARA 713
 Db 654 KITLLNKNAS-----IKLCLFSAKFR-PTNQNNQVAIVNYITIDEDQFSSRVISRG 704
 Qy 714 AFGSGQRLSPRRLRISGVNVTCEQLHFHLDTSYLRPVALVTVPALDNTTKPG--PVL 771
 Db 705 LFKENNERCLQKTMIVSQAQ-RCSEYIIHQEPSDIIISPLNLCMNISLEN--PGTNPAL 760
 Qy 772 NEGSPTSIQLVPSKCGPDNECVTDIVLQVNMIDIGSRKAPFVVRGGRKVLVSTLE 831
 Db 761 EAYSETVKVSIPIPHKDCGDDGVCSIDLVLNV--QQLPATQQQPFIVSNQNKRLTFSVOLK 819

Qy 832 NRKENAYNTSLSIIFSRNLHLASLTTPQRESPI---KVEC-AAPSAHARLCSVGHVPFQTG 887
 Db 820 NKESAYNTEIVDFSENLFFASWS-----MPVDGTEVTQIASSQSKSVTCNVGYPALKSK 875
 Qy 888 AKVTFLLEFEFCSSLLSQVFGKLTASSDSLERNGTLQENTAQTSAYIQYEPHLLFSSS 947
 Db 876 QQVTFITNFDNLQNLQNASISFRALSSEQEN--MADSNVNLKLSLLYDABIHITRST 933
 Qy 948 TLHRYE-----VHPYGTILPVGPGPEFKTTLRVQNLGCVVSGLLIISALLPAVAG 997
 Db 934 NINFYVSLDGNVSSVHVSFEDI---GPKFIFSIKV-TTGSVPVSMASVILIHIPQYTKD 988
 Qy 998 GNYFLSLSQVITNNA---SCIVQNLTEPPGPPVHP-----INPLKIGOTSSSVFSPKSENFHRELNC 1040
 Db 989 KNPLMYLTGVHTDOAGDISCEAE-----INPLKIGOTSSSVFSPKSENFHRELNC 1039
 Qy 1041 SNTQCOVVRCHLQOLAKGTEVSVGLLRVLHNEFFRRAKPSKLTVTSTFELGTBEGSVLQ 1100
 Db 1040 RTASCSNIMCWLRLDLQVKG EYFLNVSTRWNGTFAASTFTQVLTAAAEIDTYNPQIYVI 1099
 Qy 1101 TEASRSESLEVVQTRP---ILISLWILIGSVLGGILLALLVFCMLKLGFAHK----- 1153
 Db 1100 EE-----NTVTIPTIMKPEKVEVPTGVIVGSVIAGILLALLVAILMLKLGFPKRKYEM 1155
 Qy 1154 -KIPPEEKREKL 1165
 Db 1156 AKNPDETTEL 1168

RESULT 5
 S44142
 VLA-2 protein homolog - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
 R;Edelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latzer;
 submitted to the EMBL Data Library, January 1994
 A;Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not
 A;Reference number: S44142
 A;Accession: S44142
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-1178 <EDS>
 A;Cross-References: UNIPROT:Q62469; EMBL:Z39987; NID:9473098; PIDN:CAA82877.1; PID:9473098
 F;169-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 28.6%; Score 1744; DB 2; Length 1178;
 Best Local Similarity 35.3%; Pred. No. 2.6e-121;
 Matches 420; Conservative 221; Mismatches 479; Indels 70; Gaps 31;

Qy 9 LFLPLVPLT-GL--CSPFNLDHHPRLFPQPPAEFGYSVLQHVGGQRMWLVGAPWDGP 65
 Db 10 LLLQLMLVQGLNCLAYNVGLPGAKIFSGPSSBQFGYSVQQLTNPQGNMLLVGSPWSGF 69
 Qy 66 SGDRRGDYYRCPVGGAHNAPCAKGLH-GDYQLGNSSHPAVNMHGLMSLLETDDGGFPMAC 124
 Db 70 PENRMGDVYKCPV-DLPATACEKLNQNSASISNVTEIKTNMSLGLTLTENPGTGGFLTC 128
 Qy 125 APLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMYDVIIVLDGNSIYPMSEV 184
 Db 129 GPLWAHQCGNQYATGTCSDVSPDFQRTSFSAPVACPSLVDVVVVCDSNSIYPMSEV 188
 Qy 185 QTLRLRLVGLKFLIDPEQIVGLVQYGESPVHESLGDFTKBEVYVRAAKNLSRREGRETK 244
 Db 189 KNFLKPVVTGLDIPGKKTQVALIQYANEPRIIIFNLNDFETKEDMVOATSETROHGGDLTN 248
 Qy 245 TAQAIWVACTEGFSQSGHGRPEARLLVVVTDGESHGDBELPAALKACAGRVTRYGIIV 304
 Db 249 TFAIEAFARDYASQTSGGPFGATKVMVVVTDGESHGSKLKTIVIQCNDDDEILRFGIIV 308
 Qy 305 LGHYLRQRDPSPFLRIRTIASDPDRFFNFVTDAAALTDIVDALGDRIPLGEGSHAEN 364

Db 309 LGYLNRLDNTKLLKEIKAIASPTPTERYFFNVADENALLEKAGTIGEQIFSTEGT-VQG 367
 Qy 365 ESSFGLSEMSIGSTHRL--KQGLFGWGMAYDWGSGVSLWLEGHR--LPPPRMALEDEF 420
 Db 368 GDNFQEMAGVGSADYAPQNDILMLGAVGAFDWSGLV-QETSHKPVIFP-----KQAF 421
 Qy 421 PPALQ--NHAAYLYGSVSSMLLGGRRFLFSGAPRFRHKGKVIATFQKKQGVAVQAQSLQ 478
 Db 422 DQVLQDRNHSSFLGYSVAALISTEDGVH-FVAGAPRANYTQGVLYSVNKGQNVTVIQSHR 480
 Qy 479 GEIGSYFGSELCPDPTDRDGTDTDLVLAAPMELGPONKETGRVYVVLVQOQSLLTLQGT 538
 Db 481 GDIQSYFGSVLCSVDVDKTITDVLVGAFTYNDLKEEGKVLFTYTKGILNHQOFL 540
 Qy 539 LQPEPPQDARFGFAMGALPDINOGFADVAVGAPLEDHGHQALYLYHGTGSGVRPHPAQR 598
 Db 541 EGPEGTGNARFGSAIALSINDMGDFNDVIGSPVENENSGAVIYNHGQGTIRTKYSOK 600
 Qy 599 IAAA--SMPHALSYFGSVLGRDLQDGLDLDVAVGAQGAAILLSRPPIVHLTPSLEVP 656
 Db 601 ILGNGAFRRHLOFFGRLSDGVDLNGDSITDVSIGALQVQIQLWQSOSIADVAIEALFTP 660
 Qy 657 QATISVQDRCRRRQGEAVCLITAAALCFQVTSRTGRMDHOFYMRFTASLDEWTAGARAADF 716
 Db 661 DKITLKNDAK-----ITLKLCPFAERFPAQO--NNQVAILFNMTLDADGSSRVTSR 711
 Qy 717 GSGQRLSPRLR--LSGVNV--TCEQLHFHVLDTSDYLRPVALTVTFALDNTXPG--PVL 771
 Db 712 GVFRENSERFLQNMVNVQKCEHHSISIQKPSDVMVPLDLKRVDSLEN--PGTSPAL 768
 Qy 772 NECSPTSIQKLVFSDKCGDNECVTLVLQVNDIRGSKAPFVVRGGRKVLVSTLE 831
 Db 769 EAYSETVKVFSIPFYKECGSDGICISDLIDV-QQLPAIQTSQFIVSNQKRLTFFSVLK 827
 Qy 832 NRKENAVNTSLSIIFSNLHLASITPQRESPI---KYECAAPSAHARL-CSVGHVPVQTC 887
 Db 828 NRGESAVNTVLAEFSNLFPAFS-----MPVDGTEVTCVSGSKSVTCDVGFPALKSE 883
 Qy 888 AKVTFLLEPFSCSSLLSVOFGKLTASSDSLSLERNGLTQENTAQTSAYIOYEPHLLFSSES 947
 Db 884 QVTFITNFNLQNLQAAINFAQFSESQETNKA--DNSVSLTIPLLYDAELHLTRST 941
 Qy 948 TLHYEVHPGTLP-----VGPQPEFKTLRVQNLGCYVVSGLIISALLPVAHAGNYF 1001
 Db 942 NINFYEISSDENAFSVIKVEDIGPKFIFSLKV-TAGSAPVSMALVTIHIPOYTKENPL 1000
 Qy 1002 LSLSOVITNNA---SCI--VQNLTEP---PGPPVHPRELOHTNRLNGSNTQCVVRCHLG 1053
 Db 1001 LYLTGQTQAGDISCTABEINPLKPLHTAFSVSFENFRHTKELDCRTTSCSNITCWLK 1060
 Qy 1054 QLAGTEVSVGLLRVHNEFFRRAKFKSLTVVSTFELGTEEGSVLQLTEASRWSESLELV 1113
 Db 1061 DLHMAEYFVNTVTRVMNRTPAASTFTQVLTAAAEIDTHNPQ-LFVIEENAVTIPLMIM 1119
 Qy 1114 VQTPILISWILGSLVGLLLALLVFLCWKLGF--AHKKI---PEE 1158
 Db 1120 KPTEKAEVPTGVIIIGSIAGILLLLAMTAGLWKLGFPRQYKMGQNPDE 1169
 RESULT 6
 S03308
 cell surface glycoprotein cdlla precursor - human
 N;Alternate names: leukocyte adhesion glycoprotein LFA-1 alpha chain; leukocyte function
 C;Species: Homo sapiens (man)
 C;Date: 28-Feb-1990 #sequence revision 28-Feb-1990 #text_change 09-Jul-2004
 C;Accession: S03308; A47458; A47565; A48759; S36044
 R;Larson, R.S.; Corbi, A.L.; Berman, L.; Springer, T.
 J. Cell Biol. 108, 703-712, 1989
 A;Title: Primary structure of the leukocyte function-associated molecule-1 alpha subunit
 A;Reference number: S03308; MUID:89139587; PMID:2537322
 A;Accession: S03308
 A;Molecule type: mRNA
 A;Residues: 1-1170 <LAR>

A;Cross-references: UNIPROT:P20701; UNIPROT:Q9UBC8; EMBL:Y00796; NID:g31421; PIDN:CAA68
 A;Note: part of this sequence was confirmed by protein sequencing
 R;Cornwell, R.D.; Gollahon, K.A.; Hickstein, D.D.
 Proc. Natl. Acad. Sci. U.S.A. 90, 4221-4225, 1993
 A;Title: Description of the leukocyte function-associated antigen 1 (LFA-1 or CD11a) prc
 A;Reference number: A47458; MUID:93248261; PMID:8097887
 A;Accession: A47458
 A;Molecule type: DNA
 A;Residues: 1-20 <COR>
 A;Note: sequence extracted from NCBI backbone (NCBIN:130862, NCBIPI:130863)
 R;Shelley, C.S.; Farokhzad, O.C.; Arnaout, M.A.
 Proc. Natl. Acad. Sci. U.S.A. 90, 5364-5368, 1993
 A;Title: Identification of cell-specific and developmentally regulated nuclear factors t
 A;Reference number: A47565; MUID:93281759; PMID:8099450
 A;Accession: A47565
 A;Molecule type: DNA
 A;Residues: 1-20 <SHE>
 A;Cross-references: GB:M95609
 R;Nueda, A.; Lopez-Cabrera, M.; Vara, A.; Corbi, A.L.
 J. Biol. Chem. 268, 19305-19311, 1993
 A;Title: Characterization of the CD11a (alphaL, LFA-1alpha) integrin gene promoter.
 A;Reference number: A48759; MUID:93374910; PMID:8103515
 A;Accession: A48759
 A;Molecule type: DNA
 A;Residues: 1-20 <NUE>
 A;Cross-references: EMBL:Z22804; NID:g311405; PIDN:CAA80461.1; PID:g311406
 C;Genetics:
 A;Gene: GDB:ITGAL; CD11A
 A;Cross-references: GDB:119757; OMIM:153370
 A;Map position: 16p11.2-16p11.2
 C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
 C;Keywords: cell adhesion; cytoskeleton; glycoprotein; heterodimer; surface antigen; tra
 F;1-25/Domain: signal sequence #status predicted <SIG>
 F;26-1170/Product: leukocyte adhesion glycoprotein LFA-1 alpha chain #status predicted <
 F;154-317/Domain: von Willebrand factor type A repeat homology <VMA2>
 Query Match 18.1%; Score 1105; DB 2; Length 1170;
 Best Local Similarity 29.4%; Pred. No. 1.2e-73;
 Matches 366; Conservative 198; Mismatches 465; Indels 216; Gaps 55;
 Qy 11 LPLVFLTGL-----CGPFLNDEHHPRLFPGPPEA--EFGYSLVQLQHVGGQRMVLVGPWD 63
 Db 9 MAMALLSGFFFPFAPASYNLDVVGARSP-SPPGRHFGYRVLQ-VGNG---VIVGAPGE 63
 Qy 64 GPSGDRRGDVYRCPVCGAHNAPCAKHLGDYQIQLGNSHPAVNMHLGMSLLETGDDGFMA 123
 Db 64 GNS---TGSLYQCOSGTGHCLPVT-----LRGSNY--TSKYLGLTATDPTDGSILA 110
 Qy 124 CAPLWSRACSSVFSFSGICARVDASFO-PQGSILAPTAQRC-PTYMDVIVLDGNSIYP- 180
 Db 111 CDFGLSRTCDQNTYLSGLCYLFRQLQGMPLQGRPGQECIKGNVDLVFLFDGSMSLQPD 170
 Qy 181 -WSEVQTFLLRLVKGLFIDPEQIQVGLVQYGESVPVHWSLGDG---RTKEVVRVAAKNLS 236
 Db 171 EFQKILDFMDVMKKL--SNTSQFAAVQFSTSYKTEFDFSDYVVKWDPDALLKHVRHML 228
 Qy 237 RREGREFTKTAQAINVACTEGFSQSHGCRPEAARLLVYVTVTDGESHGDELPALAKACEAGR 296
 Db 229 LL-----TNTFGAINYVATEVFEELGARPDATKVLIIITDGEATDSGNIADAKD----- 278
 Qy 297 VTRYGIAVLGHYLRQRDPSFLREIRTIASDDPDERFFNVTDDEAALTDIVDLGDRIFG 356
 Db 279 IIRYIIIGIGHFQTKSQEET-----LHKFASKPASEFVKILDTPEKLDLTFELQKRIYV 333
 Qy 357 LEGSHAENESSFGLSEMSIGFSTHRLKDGILFGVMVGAYDWGSGVSLWLEGHRLPFRMAL 416
 Db 334 IEGTSKODLTSFNMELSSSGISADLSRGHAVVGAVGANDWAGGFLDLKADLQ----- 385
 Qy 417 EDEF-----PPALQNHAAAYLYGSVSSMLLGRGRRLLFLSCAPFRHHRGKVIATFQKKDGA-V 471
 Db 386 DDTFIGNLEPTPEVRAGYLYGVTVTWLPSRQKTSLLASGAPRYQHMGRVLLFQEPQGGH 445
 Qy 472 RVAQSLQGEIGSYFGSELCPDPTDRDGTDTDLVLAAPMELGPONKETGRVYVY---LVG 528

```
Db 446 SQVQTHGTQIGSYFGELCGVDVDDQGETELLIGAPFYGEQ--RGGRVFYQRRQLG 503
QY 529 QOSLLTLTQGLTQEPPODARFGPAMGALPDLDNDGADVAVGAPLEDHGHGALVYHGTQ 588
Db 504 FEEVSELQG--DGOYPL-GRFGEAITALTIDINGDLVDVAVGAPLE--QNAVYIFNGRH 558
QY 589 SGVRPHPAQIAAASMPHALSYFGRSVDGRDLDDGDDLDVAVAGAAIILLSSRPVHL 648
Db 559 GGLSPQPSQRLEGQVLSGIQWFGRSIHGVKDLGEGDGLADVAGAESQMVLSRRPVDM 618
QY 649 TPSLEVTPOAISVVQRDC---RRRQGEAVCLTAALCFQVTSRTPGRWDHOFYMR----- 699
Db 619 VTLMSFPAEIPVHEVECSYSTSNMKEGVNIT--ICFOIKSLYP-----QFQRLVANL 671
QY 700 -FTASLDEWTAGARAAADGSGQRLSPRLRLSVGNVTCEQLHFFHV-LDTSYLRPVALTV 757
Db 672 TYTLQLDGHTRRRGLFPQGRHEL--RRNIATVTSMSCTDPSFHPVVCVQDLISPINVSL 729
QY 758 TFAL---DNTTKGPNVLESGPTSIQKL-----VPSKDCDGDNECVTDLDLVQVN-MD 806
Db 730 NPSLWEEGTPROORAQKQDIPILRPSLHSETWEIPEKXGCDKKCEANLRVSPSPAR 789
QY 807 IRGRKAPFVVRGRRKVLVSTTLNKRKENAYNTSLSIIFSRNHLASLTPQR-ESPIKV 865
Db 790 SRALRLTAPF-----SLSVLSLSNLEEDAYVQLDLHFPPLGSLFRKVEMLKPHSQIPV 843
QY 866 ECAAPSAHRL-----CSVGHVPQOTGAKVTFLLBEPFSCSLLSQVFGKLTASS--DS 917
Db 844 SCBELPEERLLSALSCNVSSPIFKAGHSA-----LQMFNTLVNSSWGDS 891
QY 918 LERNGT-----LQENTAQTSAIYQIEPHLLF--SSESTLHRYEHPYGTLPVGP 965
Db 892 VELHANVTCTNEDSLEDNSATTIIPILPINILIQOEDSTLY-----VSFTPKGP- 944
QY 966 PEFTTLRVNLCGVVSGLIISALLPAVAGHGNFYLSLQVITNNASCIVQNITEPP-- 1023
Db 945 -----KIHQKHYQVRIQPSI-HDHN-----IPTLEA---VVGVPQPPSE 981
QY 1024 GP-----PVHPELQHTNRLNGSNTQC---QVVRCHGLQAKGTEVSUGLLR 1067
Db 982 GPITHQMSVQMEPPVPCYEDLE---RLPDAAEPCLPALFRCPVPRQELVQVIGTLE 1038
QY 1068 LV-----HNEPFRRAKSLTAVST--FELGTGEGSVLQLTEASRWSLSLEVVQTRPILI 1121
Db 1039 LVGEIEASSMFLSCSSLSISFNSSKHPHYGNSNLAQVV-----MKVDVYVVKQMLY 1091
QY 1122 SLMTILGSLVGLLLLLALLVFCMLKGLFFAHKKIPEBEKREKLE 1166
Db 1092 -LVVLSG--IGLLLLLLIFVLVYKVGFF-----KRNILKERME 1126

RESULT 7
I56126
Lymphocyte fuction-associated molecule-1-alpha - mouse
C/Species: Mus musculus (house mouse)
C/Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C/Accession: I56126
R/Kaufmann, Y.; Tseng, E.; Springer, T.A.
J. Immunol. 147, 369-374, 1991
A/Title: Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit
A/Reference number: I56126; MUID:91268576; PMID:2051027
A/Accession: I56126
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1163 <RES>
A/Cross-references: UNIPROT:P24063; GB:M60778; NID:g198785; PIDN:AAA39426.1; PID:g198786
C/Genetics:
C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom
F/151-315/Domain: von Willebrand factor type A repeat homology <WVAL>
Query Match 19.0%; Score 1097; DB 2; Length 1163;
```

```
Best Local Similarity 28.4%; Pred. No. 4.8e-73;
Matches 356; Conservative 192; Mismatches 470; Indels 234; Gaps 48;
QY 12 PLVFLTGL-----CSPENLDEHHPRLFPQGPPEAFGVSVLQHVGGQRWMLVGPAPMDGPS 66
Db 8 PRULLLGLQLFAKAMSYNLDTRTQSFQAQGRHFGYQLQIEDG---VVVGAPEB--- 60
QY 67 GDRRGDVYRCPVCGAHNAPCAKGLHGDYQLGNSSHPAVNMHLGMSLLETDDGCGPMACAP 126
Db 61 GDNLTGGLYHCRTSSSEFCQPVSV-----LHGSNH--TSKYLGMTLATDAAKGSLAACDP 110
QY 127 LMRACGSSVFSGGICARVDASFPQGSGLAPTAQRCTPY-----MDVVIVLDGNSI- 178
Db 111 GLSRTCDQNTYLSGLC-----YLPQSLRGPMQLQNRPAQCECMKGKVDLVFLFDGQSOLD 165
QY 179 -YPMSEVQVFLRLVGLKFLIDPEQIOVGLVOYGESPVHEMSLGDERTKEBVRVAAKNLSR 237
Db 166 RKDFEKILEPMKDVMRKL--SNTSYQFAAVQSTDCRTEFTFLDY-----VKQNKRPDV 217
QY 238 REGRE-----TKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTDGESHGDEBPAALKA 291
Db 218 LLGSVQPMFLTLTTPRINVVVAHVFKESGARPDATKVLVITDGEASDKGNISA--- 273
QY 292 CEAGRVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFPFNVTDEAALTDIYDALG 351
Db 274 --AHDITRYIIGIKHFVSQKQ-----KTLHIFASEPVEEFVKILDTFEKLDLEFTDLQ 326
QY 352 DRIFGLEGSHAENESSFGLEMSQIGFSTHRLKDGILPGMVAGYDMGSSVLWLSGGHRLFP 411
Db 327 RRIYALEGNTNRDLTSFNMELSSGISADLSKGHVAVGAVGAKDMAGGFLDLR----- 379
QY 412 PRMALED-----BPPALQNHAAAYLYGVSVSSMLLGGRRRLFLSGAPFRHRGKVIAF 463
Db 380 -----EDLQATPVQGEPLTSDVRGGLYGVYVAMWTSRSSRPLLAAGAPRYQHVGQVLLF 434
QY 464 QL-KDGAVRVAQSLQGEQIGSYFGSELCPDLTDRDGTDTDLVLAAPMFLGPQNKETGRV 522
Db 435 QAPEAGGRWNQTKIEGTQITGSYFGELCSVDLDQDGEAELLIGAPLFFGEQ--RGGRV 492
QY 523 YVYLVQOSSLTLTQGLTQEPPOD-ARFGPAMGALPDLDNDGADVAVGAPLEDHGHGAL 581
Db 493 FTV-QRRQSUFEMVSELQGDGPGYPLGRFGAAITALTIDINGDLTDVAVGAPLEB--QNAV 549
QY 582 YLYHGTQSGVRPHPAQIAAASMPHALSYFGRSVDGRDLDDGDDLDVAVAGAAIILL 641
Db 550 YIENGKPGGLSPQPSQRIQGAQVPGIRWFGRSIHGVKDLGGDLADVVVGAESGRVVVLS 609
QY 642 SRPIVHLTPSLEVTPOAISVVQRDC---RRRQGEAVCLTAALCFQVTSRTPGRWDHOFY 697
Db 610 SRPVDVVTLESPBEIPVHEVECSYSAREEQKHGKVLKA--CFRIKPLTP-----QFQ 662
QY 698 MR-----FTASLDEWTAGARAAADGSGQRLSPRLRLSVGNV-----TCEQLHFFHV-L 744
Db 663 GRLLANLSYTLQDGHMRMSRGLFPDGSHEL-----GNTSITPDKSCLDLFFHFPPI 714
QY 745 DTSYLRPVALTVTFALDNTTKGPNVLESGPT-----SIQLVFPFSKD 788
Db 715 CIQDLISPINVSLNLSL-----LEEETPDQKGRAMQPILRPSIHTVTKIPEKN 766
QY 789 CGPDNCEVTDLVLOVNMDIRGSKAPFVVRGGRKRVLSTTLNKRKENAYNTSLSIIFSR 848
Db 767 CGEDKKECANLTLS-----SPARSGPLRLMSS--ASLAVETWLSNSEDAYVWRLDLDFPR 820
QY 849 NTHLASLTP-QRESPIKVECAAPSAHARL-----CSVGHVPVQTCQAKVTFLLBEPFSCS 901
Db 821 GLSFRKVEMLQPHSRMPVSCSELETEGSSLLTKLKNVSSPIFKAGQEVSV----- 870
QY 902 SLLSQVFGKLTASS--DSLERNQTL-----QENTAQTSAIYQIEPHLLFSSSESTL 949
Db 871 --LQVMTLTLNNSWEDFVELNGCTVHCENENSSLOEDNSAATHIPLVYPVNLTKQEN- 927
QY 950 HRYEHPYGTLPVGPGRPEFTTLRVQNL-----GCYVVSGLIISALLPAVAGHGNFYL 1002
```

Db 928 -----STLYISFTPKGPKTQVQHVYQVRIQPSAYVDHNMPTLEALVGPVPHSEDLI 979

Qy 1003 SLISOVITNN--ASCIVONLTPPGPPVHPEELQHTNRLNGSNTOCVV-RCHLQQLAKGT 1059

Db 980 TYTMSVQTDPLVTHCHSDLRKPSSEARQP-----CLPGVQFRCPIVFRWEILIQVTGT 1032

Qy 1060 EVSVGLRLVHNEFFRAKPKS-LTVVSTPELGTGEGSVLQLTEASRWSES-----LLEVV 1114

Db 1033 V-----ELSKIKASSTLSLSCSSVSFNSSKHFL-YGSKASAEQVLVRKDLI 1080

Qy 1115 QTRPILSLMILIGSVLGLLLALLVFLCKLGLFFFAHKIPIBEKEEKL 1166

Db 1081 HXEMUL-HVYVLSG--IGGLVLLFLFLALYKVGFF-----KRLKXWE 1122

RESULT 8

RWU1B

cell surface glycoprotein CD11b precursor [validated] - human

N;Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac-1; leukocyte integrin alpha chain; neutrophil adherence receptor alphaM chain

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004

A;Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; 152567

R;Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.

J. Biol. Chem. 263, 12403-12411, 1988

A;Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b)

A;Reference number: A31108; MUID:88315033; PMID:2457584

A;Accession: A31108

A;Molecule type: mRNA

A;Residues: 1-1153 <COR>

A;Cross-references: UNIPROT:P11215; GB:J03925; NID:g187284; PIDN:AA59544.1; PID:g307148

A;Note: part of this sequence was confirmed by protein sequencing

R;Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.

J. Cell Biol. 106, 2153-2158, 1988

A;Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mac-1

A;Reference number: A28915; MUID:88257215; PMID:2454931

A;Accession: A28915

A;Molecule type: mRNA

A;Residues: 1-499,501-965,'P',967-1153 <ARN>

A;Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:g186935; PIDN:AAA594

A;Note: the authors translated the codon TAC for residue 1129 as Thr

A;Note: part of this sequence, including the amino end of the mature protein, was confirmed by protein sequencing

R;Shelley, C.S.; Arnaout, M.A.

Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991

A;Title: The promoter of the CD11b gene directs myeloid-specific and developmentally regulated expression

A;Reference number: A41600; MUID:92073318; PMID:1693702

A;Accession: A41600

A;Molecule type: DNA

A;Residues: 1-9 <SHE>

A;Cross-references: GB:M76724; NID:g180018; PIDN:AAA58410.1; PID:g553215

R;Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.

Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988

A;Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion molecule-1

A;Reference number: A94193; MUID:88190151; PMID:2833753

A;Accession: A30892

A;Molecule type: mRNA

A;Residues: 917-1042 <AR2>

A;Cross-references: GB:M18044

R;Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989

A;Title: cDNA sequence for the alphaM subunit of the human neutrophil adherence receptor

A;Reference number: A32218; MUID:89098893; PMID:2563162

A;Accession: A32218

A;Molecule type: mRNA

A;Residues: 9-1153 <HIC>

A;Cross-references: GB:J04145; NID:g189068; PIDN:AAA59903.1; PID:g386975

A;Note: part of this sequence was confirmed by protein sequencing

R;Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.P.; Tenen, D.G.

J. Immunol. 150, 480-490, 1993

A;Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1b gene

A;Reference number: A46526; MUID:93123748; PMID:8419480

A;Accession: A46526

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-499,501-1153 <FLE>

A;Cross-references: GB:S52227; NID:g263047; PIDN:AA24821.1; PID:g263049

A;Note: the last three bases of intron 13, CAG, are included in some but not all mature transcripts

A;Note: sequence extracted from NCBI backbone (NCBI:P:121963)

R;Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.

Biochim. Biophys. Acta 874, 368-371, 1986

A;Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across species

A;Reference number: A90664; MUID:87076671; PMID:3539202

A;Accession: A26091

A;Molecule type: protein

A;Residues: 17-31 <PIE>

A;Experimental source: granulocytes

R;Pahl, H.L.; Rogmarin, A.G.; Tenen, D.G.

Blood 79, 865-870, 1992

A;Title: Characterization of the myeloid-specific CD11b promoter.

A;Reference number: 152567; MUID:92144986; PMID:1346576

A;Accession: 152567

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-9 <RES>

A;Cross-references: GB:M84477; NID:g180184; PIDN:AA51960.1; PID:g553219

C;Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1

C;Genetics:

A;Gene: GDB:ITGAM; CR3A

A;Cross-references: GDB:120599; OMIM:120980

A;Map position: 16p11.2-16p11.2

A;Note: promoter contains a GATA motif and two Sp1 consensus binding sites

C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology

C;Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag-

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>

F;17-1108/Domain: extracellular #status predicted <EXT>

F;148-318/Domain: von Willebrand factor type A repeat homology <VWA2>

F;465-473/Region: calcium/magnesium binding #status predicted

F;530-538/Region: calcium/magnesium binding #status predicted

F;593-601/Region: calcium/magnesium binding #status predicted

F;1109-1134/Domain: transmembrane #status predicted <TM>

F;1135-1153/Domain: intracellular #status predicted <INT>

F;86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding

Query Match 17.8%; Score 1085.5; DB 1; Length 1153;

Best Local Similarity 29.6%; Pred. No. 3.4e-72;

Matches 366; Conservative 196; Mismatches 492; Indels 183; Gaps 47;

Qy 11 LPLVFLTG--LCSPFNLDHHPRLFCPPPEAEFGYSVLQVHGGQRMWLCAPWDGSGD 68

Db 3 LRVLLLTALFLCHGFNLDLTENAMTFQENARG-FGQSVVQLQSGR---VVVGAPOEIVAA 58

Qy 69 RRGDVYRCVGGAHNAPCAKHLGDLQVGNSSHP-----AVNMHLGMSLLETDDGG 120

Db 59 QRGSLVQC-----DYSTG-SCFPIRLQVPEAVNMVSLGLSLAATTSPQ 101

Qy 121 FMACAPLWSRACSSVFFSGICARVDASFOQSGSLAPTAOR-CPTY-MDVVIVLDGNSI 178

Db 102 LLACGPTVHTQCSNTYVKGCLFGLFSGNLKQPKPEALRGCPQEDSIAFLIDGSGSI 161

Qy 179 YPNSVQVTFRLRLVGLKFLIDP--EQIQ-----VGLVQYGPSPVHWSLGFRTKEEVVRA 231

Db 162 IPHD-----PRRM--KEFVSTVMEQLKSKTLFSLMOYSEEFRIHFTFKFQNNPRSL 214

Qy 232 AKNLSRRREGRETQTAQIMVACTEGFSQSHGGRPEARLLVWTDGSHSGDEELPAALKA 291

Db 215 VKPIQLLGR-THATGIRKVVRELFNITNGARKNAFKILVITDGEKF-GDPLGYEDVI 272

Qy 292 CEAGR--VTRYGIADVGLHRLRRQDRSSFLUREIIRTASDPDRFFFNVTDEAALTIDVA 349

Db 273 PEADREGVIRYVIGV-GDAFRSEKS-----RQELNTIASKPPDRHVFQVNNFEALKTIONQ 327

Qy 350 LGDRIIFGLESHAENSSFGLENSQGFSTHRLKDGILLFGMVGAYDWGGSVLWLEGHRL 409

Db 328 LREKIFAIBGTQTGSSSFHEMSQEGFAAITSNGPLLTSTVSGYDWAGGVFLYTSKEKS 387
Qy 410 FPPRMALEDFPPALQNHAAAYLYGVSVMLLRGRRRLFLSGAPFRHGRGVIAFOLKKDG 469
Db 388 TFINMTVDVS-----DNMDAYLVGA--RAILNRVQSLVGLGAPRYQHIGLVAMPR-QNTG 440
Qy 470 AVRVAQSLQEQIGSYFGSELCPDTRDGTDTVLVAAPMFGLPQNKETGRVYVYL-- 527
Db 441 MWESNANVKGQTGIGAYFGASLCSVDVDSNGSTDLVLIGAPHYY--EQTRGGQVSVCLPR 498
Qy 528 GQSSLLTLOQTLPQEPQD--ARQFANGALPDLDNODGFADVAVGAPLEDHGHGALYLYHG 586
Db 499 GORARMQCDVLYGEOQPMGRFGAALTVLGDVNGDKLTDVAIGAPGEEDNRGAIVLYFHG 558
Qy 587 TQ-SGVPHPAQRTAAASMPHALSYFGKSYVDGRLLDDGDDLDVAVGAQAAGAAILLSRPI 645
Db 559 TSGSGISPSHGRIAGSLSPRLQYFGQSLGGQDLTMDGLVDLTGVAQGHVLLRSQPV 618
Qy 646 VHLTPSLEVTPOAISVVQRDCCR---RQGEAVCLTAALCFQVTSRTFGRWDH---QPYMR 699
Db 619 LRVKAIMFPNPREVARNVFECDQVVKGEAG--EVRVCLHVQKSTRDLREGQIQSVVT 676
Qy 700 FTASLDWTAGARAAFGSGORLSPRRLRSVG--NVTCEQLHFHVLDP--TSDYLRLPVALT 757
Db 677 YDLALDSGRPHSRAPFNET--KNSTRQTVGLGTQCTELKQLPNCIEDPVSVPLRL 734
Qy 758 TFDALDNTTKPG-----PVLNMGSPSTQKLVPSKDCGPNCEVTDVLVQ---VNMDIRG 809
Db 735 NFSLVGTPLSAFGNLRLPVLAEDAQRLLTALFPPEKNGCNINICQDDLISIFSPMSLDC-- 792
Qy 810 SRKAPFVVRGRRKVLVSTTLENKENAYNTSLIIFSRNL---HLASLTPQR--ESPIKV 865
Db 793 -----LVVGGPREFNVTVIRNDGEDSYRTQVTFPPFPDLISYRKVSTLQNRQSORSWRL 846
Qy 866 ECAAPSA-----HARLCSGVHPVFTQGAKVTFLEPEFSCSLLSQVQKLTASDLSL 918
Db 847 ACESASSTEVSGALKSTCSINHPFENSEVENTFIDVDSKASLG---NKLK-----L 898
Qy 919 ERNQTLOENTAQ-----SAYIQYBPH-----LLPSSSTLHRYBVHPYGT 959
Db 899 KAVTSENMPRTNKTBFQLEPLVKYAVYVVTSHGVSTKYVLTNFTAGENTSRYVMQHY-- 956
Qy 960 LPVGPGEPEKTLRVQNLGCVVWSGLIISA--LLPAVAHGNGYFLSLSQVI----- 1008
Db 957 -----QVSNLG---QRSPLSLVFLVP-----VRLNQTVIMDRPQVTF 991
Qy 1009 TNNAACVQNLTEPPGPPVHPEELQHTNRLNGSNTQCVVRCHLQGLAKGTEVSVGLLRL 1068
Db 992 SENLSSTCHTKERLPSHSDFLAELRKAPVNVNCISIAVCQRIQCDDIPFFGIQBEFNATLGN 1051
Qy 1069 VHNFFPRRAKSLTVSTPELGTREGSVLQLTEASRWSLSLAV--VQTRPILISLWILI 1127
Db 1052 LSPDWYIKTSHNHLIVSTAEILNDSVFTLLPQGGAFVRSQTKVPEFVNPPLVIV 1111
Qy 1128 GSVLGGLLALLVFLWKLGFPAHKKIPBEEKREEK 1164
Db 1112 GSSVGGLLALLALITAAVLYKLGF-----KQYK 1139

RESULT 9

RWHLIC

cell surface glycoprotein CD11c precursor - human

N/Alternate names: leukocyte adhesion receptor p150,95 alpha chain

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004

C/Accession: A36584; A35543; S00864

R/Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.

J. Biol. Chem. 265, 12750-12751, 1990

A/Reference number: A36584

A/Contents: erratum

A/Accession: A36584

A/Molecule type: DNA

A/Residues: 1-1163 <COR>

A/Cross-references: UNIPROT:P20702
R/Note: this revision to the sequence from reference A35543 includes the carboxyl end
J. Biol. Chem. 265, 2782-2788, 1990
A/Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.
A/Reference number: A35543; MUID:9015306; PMID:2303426
A/Accession: A35543
A/Molecule type: DNA
A/Residues: 1-834 <CO2>
R/Note: this sequence has been revised in reference A36584
R/Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.
EMBO J. 6, 4023-4028, 1987
A/Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte
A/Reference number: S00864; MUID:8816645; PMID:3327687
A/Accession: S00864
A/Molecule type: mRNA
A/Residues: 1-755, 'L', 757-1163 <CO3>
A/Cross-references: GB:M81695; EMBL:Y00093; NID:G487829; PIDN:AA59180.1; PID:G487830
R/Note: part of this sequence was confirmed by protein sequencing
C/Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on mye
C/Genetics:
A/Gene: GDB:ITGAX; CD11C
A/Cross-references: GDB:119758; OMIM:151510
A/Map position: 16p11.2-16p11.2
C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homol
C/Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;
F/1-19/Domain: signal sequence #status predicted <SIG>
F/20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>
F/20-1107/Domain: extracellular #status predicted <EXT>
F/149-319/Domain: von Willebrand factor type A repeat homology <VWA4>
F/1108-1133/Domain: transmembrane #status predicted <TM>
F/1134-1163/Domain: intracellular #status predicted <INT>
F/61,89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status pres
Query Match 17.3%; Score 1054; DB 1; Length 1163;
Best Local Similarity 28.5%; Pred. No. 7.8e-70;
Matches 356; Conservative 201; Mismatches 474; Indels 220; Gaps 50;
Qy 13 LVFLTGLCSF--FNLDEHHPRLFPPEAFGYSVLQHVGGQRMVLVGPMDGPGSDRR 70
Db 8 LLLFTALATSLGNLTTELTAF-RVDSAGFGSVVQVYANS---WVVGAPQKITAANQT 63
Qy 71 GDVYRCFVGGAHNAPCAKGLHDYQLGNSHP--AVNMHLGMSLLETGDGCGFMACAPLWS 129
Db 64 GGLYQC-----PIGLQVPPAVNMVSLGLSLASTTSPSOLLACGPTVH 113
Qy 130 RACGSSVFSFGICARVDASFPQPGSLAPTAQRCPTY-MDVVIVLDGNSIYP--WSEVQT 186
Db 114 HECGRNMYLTGLCFLLGPT-QLTQRLPVSRQECPRQEQDIVFLIDGSGSISSRNFATMMN 172
Qy 187 FLRLVVKLFIDPEQIQVGLVQYGESPVHWSLGDPTKEEVVRAAKNLSRRRGRETHTA 246
Db 173 FVRAVISQ-FQRP-STQFSLMQFNKFTHTFEFRRTSNPLSLASVHQLQG-FYTTA 229
Qy 247 QAIWVACTEGFSQSHGSRPEARLLVVVTDGESH-DEELPAALKACEAGRVTRYIAVL 305
Db 230 TAIQNVVHRLPHASYGARRDATKILIVITDGKSGSLDYKVIWPAADAAGIIRYAIQVG 289
Qy 306 GHVLRQRDPSSFLERITIASDPDERFPFNVVDEAALTDIVDALGDRIFGLSGSHANE 365
Db 290 LAFQNR-----NSWKELNDIAKPSQEHIPKVEDFDALQIQOLKEKIFAIGTETTS 344
Qy 366 SSFGLMSQIGFTHRLKDGILPGMVGAYDWGCVLMLEGGHRLFPFRMALEDFPPALQ 425
Db 345 SSFELEMAQEGFAVTPDQVILGAVGSGFTW-----SCGAFLYPNMS-----PTPI 391
Qy 426 NHA-----AYLYGVSVMLLRGRRRLFLSGAPFRHGRGVIAF-OLKKDGAVRVAQS 476
Db 392 NMQENVMDRDSYLYGSTELALWKVQSLV-LGAPRYQHTGKAVIFQVSRQW--RMKAE 448
Qy 477 LQGEQIGSYFGSELCPDTRDGTDTVLVAAPMFGLPQNKETGRVYV----- 524
Db 449 VTGTQIGSYFGASLCSVDVDTDGTDLVLIGAPHYY--EQTRGGQVSVCLPRGMRWMC 506

C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C/Accession: I58409; K.; Ueda, R.; Horio, Y.; Murata, Y.; Tamari, M.; Uchida, K.; Takah
R.; Hibi, K.; Yamakawa, K.; Ueda, R.; Horio, Y.; Murata, Y.; Tamari, M.; Uchida, K.; Takah
Oncogene 9, 611-619, 1994
A/Title: Aberrant upregulation of a novel integrin alpha subunit gene at 3p21.3 in small
A/Reference number: I58409; MUID:94119603; PMID:8290272
A/Accession: I58409
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1035 <RES>
A/Cross-references: UNIPROT:Q13797; GB:D25303; NID:9464180; PIDN:BAA04984.1; PID:G533327
R;Palmer, E.L.; Ruegg, C.; Ferrando, R.; Pytela, R.; Sheppard, D.
J. Cell Biol. 123, 1289-1297, 1993
A/Title: Sequence and tissue distribution of the integrin alpha9 subunit, a novel partne
A/Reference number: A49459; MUID:94064789; PMID:8245132
A/Accession: A49459
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 30-1035 <PAL>
A/Cross-references: GB:L24158
C;Superfamily: integrin alpha-4 chain
C;Keywords: glycoprotein; metal binding; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>

Query Match 12.1%, Score 738; DB 2; Length 1035;
Best Local Similarity 22.6%; Pred. No. 2.5e-46;
Matches 284; Conservative 179; Mismatches 446; Indels 350; Gaps 44;

Qy 9 LFLPLVLTGLCSFNLDEHHRLFCPPAEFGYSVLQHVGGQRMVLGAPWD----G 64
Db 16 LLUALVAGTIPAGAYNLDOPQRFVHGFADSPFFGYAVLEHFDNTRVLVVGAPKADSKYS 75
Qy 65 PSGDRRGDVYRCPVGGAHNAP-----CAKHLGDYQLGNS-SHPAVNMHLGMSLL-E 114
Db 76 PSVKSPGAVFKCRV--HTPDRRCTELDMARGKNRGTSCKTCREDRDDEWNGVSLARQ 132
Qy 115 TDGDDGPMACAPLWSRAC--GSSVFSSGICARVDASFPQPGSLAPTAQRCPTMYDVVIVL 172
Db 133 PKADGRVLACARWKNYYEADHILPHGFCYIIPSNLQAKG----- 173
Qy 173 DGSNSIYPMSVEQTFLLRLVGKLFIDPEQLQVGLVQYGSPVHWSLGDPTKEEVVRAA 232
Db 174 -----RT----- 175
Qy 233 KNLSRREGRETAKTAQAIMVACTEGFSQSHGGRPEMARLLVVVTDGSHDGEELPAALKAC 292
Db 176 -----LIFCYEYKKY-----GHEG-----SC 194
Qy 293 EAGRVTRYGIHVLGYLRRORDPSSFLREIRTIASDPDERFFNVTDAAALDIVDALGD 352
Db 195 QAG-----IA----- 199
Qy 353 RIFGLESHANESSFGLMSQIGFSTHRLKDGILFGWGCAYDWGGSVLWLEGGHRLFP 412
Db 200 -----GFFTEEL---VVMGAPGSFYWAGTIKVLNLTNDTY-- 231
Qy 413 RMALEDFPAPLQNHAAIYGVSYSSMLLRGRL-FLSGAPRFRHRKGKIAFOL-KDGA 470
Db 232 -LKLNDEV--IMRRYTYLGIATVAGHFHSHPSTIDVVGAPQDKGIGKVIYFRADRSRT 288
Qy 471 VRVAQSLQGBQIGSYFGSELCLPDLTRDGTDTDLVLAAPMFLGPNKQKTRGVVYLVGQQ 530
Db 289 LKIFQASGKKMSYFGSSLCAVDNLNGDGLSD-LLVGAPMF--SEIRDEQVTVYINRGN 345
Qy 531 SLTLQGTLOPEPPQDARFGANGALPDLNQDGFADVAVGAPLEDHGQGXLYHGTQSG 590
Db 346 GALEEQALATGDGAYNAHFGEASLASLDLNDGFPDVAIGAPKEDDPAGAVIYHGDAGG 405
Qy 591 VRPHPAQRIAAASWPHALSVEGSDVGRLLDLDGDDLVDVAVGA--QGAATLLSSRPVHL 648
Db 406 IVPOYSMKLGGKINPVLVRMFGISGIGDMDNGVDPDVTVGAFMDSVVLRLARPVITV 465
Qy 649 TPSLEVTFOAISVVQRDCRRRGOEAVCLTAAALCF-----QVTSRTP 689

Db 466 DVS I - FLPGSINTAPQCHDGOQPVNCLNVTTCSFHGKHVPBEEIGLNYVLMADVAKKEK 524
 Qy 690 GRWDHGFYMRFTASLDIEWTAGARAAFDGSGORLSPPRLRLSVGNVTCEQLHFHV-LDTSD 748
 Db 525 GQMPRVYFVLLGTFMGQVT-----EKLQLTYYEEECRRHYVAHVKKRVQD 568
 Qy 749 YLRPALVTVTFAL-----DNTRKP-GPVL--NEGSPTSIOKLVPFKDCDQPDNECVT 797
 Db 569 VISPIVFEAAYSLESHVTGBEERELPPLTVLRWKKGQKIAQKNQTVFENCRCSE-DCAA 627
 Qy 798 DLVLQVNMDIRG-SRKAPFVVRGRRKVLVSTTLENKENAYNTLSIIIFSRNLHLASLT 856
 Db 628 DLQLOQKLLSSWDEKTLXLLALGAVKNI SLNISGLDDAYDANVSFVNSRELFFINMW 687
 Qy 857 PQRESPIKVECAAPSAAHARLCSGVHPVOTGAKVTFLLPEFSCSSLLSQVFGKL-TASS 915
 Db 688 QKEE--MGISCELLESDFLKCSGVFFPMRSKYSKVESVIFDTHLSGEEVLSFIVTAQS 745
 Qy 916 DSLERNGTLOENTAQTSAYIQYE-----PHLLFSSEST-----LHRYEVHPYG 958
 Db 746 GNTRESLHDNTLVLMLVPLMHEVDTSITGIMSPTSFVYGESVDAAFIQLDDLECH--- 802
 Qy 959 TLPVGPQPEKTTLRVQNLGCYVVSGLIISALLP-AVAGCGNVFELSQVIT--NNASCI 1015
 Db 803 FQPI-----NITLQVYNTGPTSLPGSSVISFPNRLSSGGAEMFHVQEMVVGQEKNGCS 856
 Qy 1016 VQNLTPBPQPVHPBELOHT-----NRLNGSNTQCQVVRCHLQGLAKGTEVSU 1063
 Db 857 FQKNPTPCIPOQENIFHTIFAFYTKSGRKVLDCCKPGIGISCULTAHCFNSALAKEBSRTI 916
 Qy 1064 GLLRLVHNE-----FFRRAKFK-----SLTVVSTFELGTSEGSVLQUTASRWSES 1109
 Db 917 DIYMLNLTETLKKDSSSVIQFMSRAKVQVDPALRVVLAHGNFEVTV--VFEA----- 968
 Qy 1110 LLEVQTRPILISLWILIGSVLGGLLALLVLCWLKLGFF--AHKIPBEEKREEKLE 1166
 Db 969 -LHNLEPRGVVVG-WIAAISLLVGILLFLLAVLLWMKGFFRRRYKEIIEAKNRKENE 1025

RESULT 13
 JC7294
 alpha integrin - sea urchin (Strongylocentrotus purpuratus)
 C/Species: Strongylocentrotus purpuratus (purple urchin)
 C/Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004
 C/Accession: JC7294
 R/Susan, J.M.; Just, M.L.; Lennarz, W.J.
 Biochem. Biophys. Res. Commun. 272, 929-935, 2000
 A/Title: Cloning and characterization of alpha integrin in embryos of the sea
 A/Reference number: JC7294
 A/Contents: Embryo
 A/Accession: JC7294
 A/Molecule type: mRNA
 A/Residues: 1-1054 <SUS>
 A/Cross-references: UNIPROT:Q9U6S1; GB:AA55724
 C/Genetics:
 A/Gene: snalphpap
 C/Superfamily: integrin alpha-2b chain
 C/Keywords: calcium binding; embryo; glycolysis; glycoprotein; heterodimer; tra

Query Match 12.0%; Score 732.5; DB 2; Length 1054;
 Best Local Similarity 23.4%; Pred. No. 6.5e-46;
 Matches 294; Conservative 188; Mismatches 432; Indels 345; Gaps 47;

Qy 10 FLPLVFLTGL-----CSPFNLDHHRLRPGPPEAEFGYSVLQHVGGQRWMLVG---AP 61
 Db 4 FLLLSFTTLVLLDSTVAFNFLRAPVKFDPGQGSILFGFSAQHRDQNTDWMVLIGAPEAP 63
 Qy 62 WDGPSPGRRGDVYRCPVG- AHNAPCAK---GHLGDYQLGNSHPAVNMHLGMSLLETGD 117
 Db 64 TTQPGVTNGAVYKCPVTPUSGGGCPCEQVFDTTGNTVELDKS-----NQWFGATLASSGP 119
 Qy 118 DGGFMACAP--LWSRACSGSVFSSGICARVDASFOQGSILAPTAQRCPVTYMDVIVLDGS 175

Db 120 DGRILACAPRLVLTQSTISPT-----141
QY 176 NSIYPMSEVQTFLLRLVGKLFIDPEQIQVGLVQGVSPVHWSLGDPRTKKEVVRAAKNL 235
Db 142 -----141
QY 236 SRREGRETKQAIAWVACTEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALAKACEAG 295
Db 142 -----DKEREPT-----148
QY 296 RVTRYGIAVLGHYLRORDSSFLREIRTTASDPDERFFNVNTEAALTDIVDALGRIF 355
Db 149 -----GTCFVGH-----SDFTNFVYSPCQSTDRD-----LF 175
QY 356 GLEGSHAENESSFGLEMSQIGFSPHRLKDG--ILFGMVGAYDMGGSVLMLEGGHRLPPPR 413
Db 176 GFD-----KITHCQAGFSAQIPSDNSTLVMGAPGSI-----YLQG--QIFAQ 216
QY 414 MA-LED-----EFPALQNHAAVLGYGVSSMLLGG--GRRFLSGAPRFRH--RGKVIAPQL 465
Db 217 LSTLSDVSNTPQAVAFDN--SYRGYSALGDFNGDGLDYVVGTPRGESLRLGLVAIFD- 273
QY 466 KKGAVRVAOSLQEQIGSFVSGELCPDTRDGTDTVLLVAAPMFL-----GPNKETGR 521
Db 274 --OSLVEIITPVVGEQIVSYFGYSVASVDVNGDGLDD--LLVGAPMFTNREPATEKWEAGR 330
QY 522 VYVYL-----VGOOSLLTLQGTLOPEPPQDARFGFANGALPOLNODGFADVAVGPLE- 574
Db 331 VYVYLNADHSLGAPQLTKKI-----RARPGFPTISGDSNQDGFNDVAIGAIDYG 383
QY 575 DGHGALYLYHGTSQGVRRPAPQRIAAAASMPHA--LSYFGRSVDGRLLDDGDLDVDAVGA 633
Db 384 EDNSGVVYVHGSAGELRLTESQVLTPELGSFSDITTFGFSVDGGQDMQNDYDPLVGA 443
QY 634 QG--AAILLSRPIVHTPSLEVTPQAISVVQRDCR--RRGOEAVCLTAALCFQVNT-SRTP 689
Db 444 ESADAAILVRTPVVLLEAEITIEPIGINDKNKYELPDGTMVTSFVAMACFTVTGNHLP 503
QY 690 GRWDHQFVMTASLDEWTAGARAAPFGSGORLSPRRLSLVGNVTCEQLHFFVLDT-SD 748
Db 504 ARIGISVTLTVDSI-----TSGRALLVNLSELQVTKNRNLDVGMKFCDDLRAVYVNTIQD 560
QY 749 YLRPVALTVPFALDNTT-----KGPVLNBSPTSIOKLVPFSCDGPDPNECVTDLVQV 803
Db 561 KLTPIAVDLQVELTDESILLPYELPIINKEAVSSQTKQVSIQNNC--VNNICIPBIGITV 619
QY 804 NMDIRGRKAPFVVRGRRKVLVSTTLLENKENAYNTSLIIFSRNLHLASLTTPQRESPI 863
Db 620 TPNL-----PNVIGQAQLTLVVSINNRRGEDAFQSLAVYYPEGLQYVRLERRANMDF 673
QY 864 KVECAAPSAHARL--CSVGHVPVQFGAKVTPLEPEFSCSLLSQVFGK-----LTASS 915
Db 674 SVTCTEDSALRMITCDGNPLVG-----KYNLEFGLTLSTL--QVSGDKONIEPYLVAGS 726
QY 916 DSLERNGTQLONTAQTSAYIQYEP--HLLFSSESTLHRYEHPYGTLPVGP-----964
Db 727 ENNEDPNTLNNELNVTVAVIVDATKLLSASYPEIVTVRVPEDNIVPEFPTKNASEADI 786
QY 965 GPEFKTTLRVQNLGCYVVGSLIISALLPAVAHGNYFLSLISQVITN--NASC-IVQNLTPE 1022
Db 787 GMEVHVLYEVRNTGSSNAAEVTLNIRWPEKDNDYLLPYLLGIMTDEGVTQCISQGANP 846
QY 1023 PGPPVHPEELQHTNRLNGSNQ-----CQVVRCHLGQL 1055
Db 847 LGVKL-----EASTKEQLSNSTTVSGRRKRGEYAEALAAQAEPICTPESCVLINCTIDEI 903
QY 1056 AKGTEVSVGLRLVHNBFFRAKFKSLTVVSTFELTEEGSV-----LQLTBASR 1105
Db 904 KATSKVRIILGRFWERTFQKAVSEAVPVQIQTATATATVRSIPIYNIPLMFEFTDSTK 963
QY 1106 WSESL--LEVVTQRPILSLMILIGSLGGLLLALLVFLWKLQGFPAHKKIPESKEEBE 1163

Db 964 ASTLITABELVLPVSVSIAMWIIIVSVLGGIILLIILGLMKGPFERKKPGEDQKEYE 1022
RESULT 14
A41131
lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse
N;Alternate names: integrin alpha-4
C;Species: Mus musculus (house mouse)
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 09-Jul-2004
C;Accession: A41131; S16742
R;Neuhaus, H.; Hu, M.C.T.; Hemler, M.E.; Takada, Y.; Holzmman, B.; Weissman, I.L.
J. Cell Biol. 115, 1149-1158, 1991
A;Title: Cloning and expression of cDNAs for the alpha subunit of the murine lymphocyte-
A;Reference number: A41131; MUID:92064645; PMID:1840602
A;Accession: A41131
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1039 <NU>
A;Cross-references: UNIPROT:Q00651; EMBL:X53176; NID:G51484; PIDN:CAA37316.1; PID:G51485
C;Superfamily: integrin alpha-4 chain
C;Keywords: cytoskeleton; transmembrane protein
Query Match 11.3%; Score 690; DB 2; Length 1039;
Best Local Similarity 22.7%; Pred. No. 9.4e-43;
Matches 286; Conservative 186; Mismatches 396; Indels 392; Gaps 59;
QY 23 FNLDEHHPLRFGPPEAEFGYSVLQHVGGQRWMLVGAP---W-DGPGDGRDDGVYRCVP 78
Db 41 YNLDPENALLYQGPSGTLFGYVVLHSHGSKWLVGAPTASWLSNASVVPQAIYRCGI 100
QY 79 -----GGAHNAPCAKGHLDGYQLGNSSHPAVNMHLGMSLLETGDP--GGFMAC 124
Db 101 RKNPNQTCBQLQSGSPSGPCGKTCLEERD-----NQLGVTLRQPGENGSIIVTC 151
QY 125 APIWSRACSSVPSSGICARVDASFPQCSLAPTAQRCPTYMDVIVLDGNSNIYPSSEV 184
Db 152 GHRW-----KNITF-----YMKSDNKUPT-----169
QY 185 QTLRLRLVGLKFTDPQIQVGLVQGVSPVHWSLGDPRTKKEVVRAAKNLSRREGRETK 244
Db 170 -----GICYMP-----SDLRTE-----LSKR-----186
QY 245 TAOAIMVACTEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALAKACRAGRVTTRYGIAV 304
Db 187 -----MAPCYKDVTRKFGEN-----FASCQAG-----208
QY 305 LGHYLRQRDPSPFLREIRTTIASDPDER-----FFPNVTDEAALTDIVDALGRIFGLE 358
Db 209 -----ISSFYTDLIVMGAPGSSYMTGTVFYVYNTIT-----TNQYKAFYDR-----248
QY 359 GSHAENESSFGLEMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGGHRLPPRMALED 418
Db 249 -----QNVQVFG-----255
QY 419 EFPALQNHAAVLGYGVSSMLLGGRRRL--FLSGAPRFRHGKVIAPQLKKGAVRVAQSL 477
Db 256 -----SYLGYSVGAHGRPSPTTEVVGAPQHEQIGKAYIFSI--DENELNIVYEM 304
QY 478 QGEQIGSYFGSELCPDTRDGTDTVLLVAAPMFLGPQN--KETGRVYVYLVQQS--LL 533
Db 305 KGKGLASVYFGASVCAVDLNAFGSD--LLVGAPM-----QSTIREGRVYVYVINGMGAMV 359
QY 534 TLOGTLQPEPPQDARFGFANGALPDNLNODGFADVAVGAPLEDGHQCALYLYHGTQSGVRP 593
Db 360 EMBRLVSGDKYAAARFGESIANLGDINDGDFEDIAFGAQEDDLRGAIVYINRGVGISS 419
QY 594 HPAQRIAAASMPHALSYFGRSVDGRLLDDGDLDVDAVGA--QGAAILLSSSRPIVHLTPS 651
Db 420 TYSORIEGQQIKSLRMFGQSIGQIDADNNGVVDVAVGAFQSDSAVLLKTRPVPVIVEAS 479
QY 652 LEVTPQAISVVQRDCRRRQGEAVCLTAALCFQVTSR--TPGRWDHQFYMFMFTASLDE----706
Db 480 LS-HPESVNRKTPDCTENGLPSVCMHLTLCFSYKGEVPG-----YIVLFVNVSLDVHRKA 534

QY 707 -----WTAGARAFDGGORLSPRLRLSVGNVTCEQLH--FHVLDTSVLRPVALT 756
 Db 535 ESESRFFYFNGSDVITGS-----IRVSSGKCR-RTHQAFMRKQVDRDLTPHIVE 585
 QY 757 VTPALON--TTKPG-----PVLNKGSPS-IQKLVPSKCGPDNECVTDLVLQVNM 805
 Db 586 AYHLGHVHTKNTTEFFPPLQLOKKEKQVIRKMINFARFCAYEN-CSADLQVSAKV 644
 QY 806 D-IRGSRKAPFVVRGGRKVLSTLTLENKENAYNTSLSIIFSRNLHLASLTPORESPIK 864
 Db 645 GFLKPYENKTYLAVGSMKTMNLVSLFNAGDDAYETTLNQLPTGLYFIKILDEEKOIN 704
 QY 865 VECAPSAHARL-CSVGHVPVQCAK--VTFLEBFESCSLLSQVFGKLTAS-SDSLER 920
 Db 705 CEYTESGIVKLACSLGIYVDRSLRIDISFLDVL-----SSLSRAHEDLSISVHASCEN 759
 QY 921 NGTL---QENTAQTSAYIQEPHLLFSSESTLHRYEYVHP-----YGTLPVGFGE----- 967
 Db 760 EGELDQVDRNRVLTITPLREYVML-----TVHGL-VNPTSFVYGSSEENE-PETCMAEK 811
 QY 968 FKTLRQNLGCVYVSGLIISALLPAVAGGNVFL-----SLSQVITNASC----- 1014
 Db 812 LNLTFHVINTGISMAPNVSVKIMVP-----NSFLPQDDKLENVLDVQTTTGOCHPKHYG 865
 QY 1015 -----IVQNLTPEPPPPVHPELOHTNR-----LNGSNVQCVVRCHLQOLAGTE 1060
 Db 866 RECTFAQKGIAGTLTD-----IVKFLSKTDKRLLYCMKADQCHDCLDFCNFGKMSGKE 919
 QY 1061 VSVGLLRVLHNEFFRRKAKSLTVST--FELGT-----EBSVLQLTASRSSESLE- 1112
 Db 920 AS-----VHIQEGRSIEMDETSLSKFEIKATAPPEHPKVIENKQENVAHVLEG 973
 QY 1113 VVQTRP-----ILISWLILGVLGGLLALLVFC-LWKLGF--AHKKIPEEEKREE 1163
 Db 974 LHQRPKRHTIIIIITISLLG-----LIVLLIISCVMMKAGFKFKQYKSIQLEENRD 1027

RESULT 15
 S06046
 integrin alpha-4 chain precursor - human
 N;Alternate names: CD49d; very late antigen-4 alpha chain; VLA-4 alpha chain
 C;Species: Homo sapiens (man)
 C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 09-Jul-2004
 R;Accession: S06046; A39355; D28018
 R;Takada, Y.; Ellices, M.J.; Crouse, C.; Hemler, M.E.
 EMBO J. 8, 1361-1368, 1989
 A;Title: The primary structure of the alpha(4) subunit of VLA-4: homology to other integ
 A;Reference number: S06046; MUID:89356603; PMID:2788572
 A;Accession: S06046
 A;Molecule type: mRNA
 A;Residues: 1-1038 <TAK>
 A;Cross-references: UNIPROT:P13612; GB:X16983; EMBL:X15356; MID:g33945; PIDN:CAA34852.1.
 R;Rosen, G.D.; Birkenmeier, T.M.; Dean, D.C.
 Proc. Natl. Acad. Sci. U.S.A. 88, 4094-4098, 1991
 A;Title: Characterization of the alpha-4 integrin gene promoter.
 A;Reference number: A39355; MUID:91239513; PMID:2034655
 A;Accession: A39355
 A;Molecule type: DNA
 A;Residues: 1-93 <ROS>
 A;Cross-references: GB:M62841
 R;Takada, Y.; Strominger, J.L.; Hemler, M.E.
 Proc. Natl. Acad. Sci. U.S.A. 84, 3239-3243, 1987
 A;Title: The very late antigen family of heterodimers is part of a superfamily of molec
 A;Reference number: A94151; MUID:87204112; PMID:3033641
 A;Accession: D28018
 A;Molecule type: protein
 A;Residues: 40-50, 'E', 52-53 <TA2>
 C;Genetics:
 A;Gene: GDB:ITGA4; CD49D
 A;Cross-references: GDB:128032; OMIM:192975
 A;Map position: 2q31-2q32
 C;Superfamily: integrin alpha-4 chain

C;Keywords: cell adhesion; cytoskeleton; duplication; signal sequence #status predicted <SIG>
 F;1-39/Domain: integrin VIA-4 alpha-4 chain #status predicted <MAT>
 F;40-1038/Product: integrin VIA-4 alpha-4 chain #status predicted <MAT>

Query Match 11.2%; Score 684; DB 2; Length 1038;
 Best Local Similarity 22.9%; Pred. No. 2.6e-42;
 Matches 289; Conservative 181; Mismatches 423; Indels 368; Gaps 54;

QY 9 LFPLVFLTGLCSFNLDEHHPLFPFPPEABFGYSLVQHVGGQRMVLGAP---W-DG 64
 Db 28 LLLCLGVPFG--RPYVNDTESALLYQCPHNTLFGYSLVLSHGANGRLLVGAPANTLAN 85
 QY 65 PSGRDGDVYRCVPV-----GAHNA-PCAKGHLDYQLGNSSHPANVHMGMS 111
 Db 86 ASVINPGAIRCRIGKNKPGQTCQLQGLSPNGPCGKTCLERD-----NQLGVT 136
 QY 112 LLETGDD-GGFMACAPLWSRACGSSVSSGICARVDASFOQGLAPTAQRCPTMDVVI 170
 Db 137 LSRQPGENGSIIVTCGRW----- 154
 QY 171 VLDGNSIYPMSEVQTFRLRLVGLFIDPEQIOVLVOYGESPVHWSLGDFTKEEVR 230
 Db 155 ---KNIFYIKNE-----NKLPTGGC-YGVPP-----DLRT----- 180
 QY 231 AAKNLSRREGRETKTAQAINVACTEGESQSHGGRPEAARLLVVVTDGESHGDELPAALK 290
 Db 181 ----- 187
 QY 291 ACBAGVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNVNDEAALTIDVDAL 350
 Db 188 PCYQDIYKFG----- 198
 QY 351 GDRIFGLEGSHAENESSFGLMSQIGFSTHRLKDGILFGMVAGYDMGGSVLWLE---GGH 407
 Db 199 -----ENFAS-----CQAGISSPYTKDLIVMGAPGSSYMTGSLFVYNITNKY 241
 QY 408 RLFPFRMALEDEFPALQNH---AAYLYGYSVSSMLLRGGRRL-FLSGAPFRIRGKVIAP 463
 Db 242 KAFLDK-----QNOVKFGSYLGYSVGAGHFRSQHTTEVVGAPQHEQIGKAYIF 290
 QY 464 QLKKGCAVRVAQLOEQSGYFSGSELCPDTRDGTDLVLLVAAMPFLGPQN--KETGR 521
 Db 291 SI-DEKELNILEMKKKLGSYFGASVCAVDLNADGFS-DLVGAPM-----QSTIREGR 344
 QY 522 VVYVL-VGQOSSL-TLQGLTQPEPPQDARFGFANGALPDNLQDGFADVAVGAPLEDGHQ 579
 Db 345 VFVYINGSGAVNMAMETNLVSGDKYAARFGESIVNLGIDINDGDFEDVAIGAQQEDDLQ 404
 QY 580 ALYLYHGTQGVPRPAPQRIATAASMPHALSYFGRSVDGRLLDGDGDLVDVAVGA--QGA 637
 Db 405 AIYIYNGRADGISSTFSQRIEGLQISKLSMFGQISGGIDADNNGYVDVAVGAFRSDSA 464
 QY 638 ILLSSRPVHLTSLVTPQAIISVQVORDCRRRQEAVALTAALCFQVTSR-TPGRWDHQF 696
 Db 465 VLLRTPVIVDASLS-HPESVNRTKFDVCVENGWPSVICIDLTLCFSYKGEVFG----YI 519
 QY 697 YMRFTASLD---EWTAGARAAFDSG-QRLSPRLRLSVGNVTCEQLH--FHVLDTSVLR 750
 Db 520 VLFYNNMSLDVNRKAESPFPFYFSSNGTSDVITGSIQVSSREANC-RTHQAFMRKQVDRIL 578
 QY 751 RPVALTVTAL-----DNTYKPGVNLNKGSPSI-OKLVPSKCGPDNECVTDL 799
 Db 579 TPIQIEAAVHLGPHVIVSKRSTEEFPPLQLOKKEKDKIMKKTINFARFAHEN-CSADL 637
 QY 800 VLQVND-IRGSRKAPVVRGGRKVLSTLTLENKENAYNTSLSIIFSRNLHLASLTPO 858
 Db 638 QVSAKIGFLPKPHENKTYLAVGSMKTMNLVSLFNAGDDAYETTLNQLPTGLYFIKIL 697
 QY 859 RESPIKVECAAPSAHARL-CSVGHVPVQ--TGAKVTFLEBFESCSLLSQVFGKLT--- 912
 Db 698 EEKQINCEVTDNSGVQLDCSIGIYVDRSLRIDISFLDVL-----SSLSRAHEDLSITV 752
 QY 913 -ASSDSLERNGLQENTAQTSAYIQEPHLLFSSESTLHRYEYVHP-----YGTLPVGFGE 967

Db 753 HATCENEEMDNLKHRSVTVAIPLKYEVL-----TVHGF-VNPTSFVYGSNDENE-PE 804
Qy 968 -----FKTTLRVONLGCYVVSGLIISALLP-AVAHGGNYFLSLSQVITNNASCIQNL 1020
Db 805 TCWVEKNLTFHVINTGNSMAPNVSVIIMPNSFSPOTDKLFNILDVQTTGEC----- 858
Qy 1021 EPPGPPVHPELQHTNRLNGSNTQCQVVR-----CHLG 1053
Db 859 -----HFENYQVCALEQQKSAMQTLKGIVRFLSKTDKRLLYCIKADPHCLNFCNFG 911
Qy 1054 QLAGTEVSVGLLVLVNEFRRRAKFKSLTVVST--PEL---GTEEGS--VLQLTEASRW 1106
Db 912 KMESGKEAS-----VHIQLEGRPSIILEMDETSALKFEIRATGFPPEPNRPRVIELNKDENV 965
Qy 1107 SESLEVV--QTRFILSLMILIGSVLGGLLALLVFCILWKLGFF--AHKKIPEREKRE 1162
Db 966 AHVLEGLHHQRPKRYFTIIVISSLLGLIVLLLSIVMMKAGFFKRYKSIQENRR 1025
Qy 1163 B 1163
Db 1026 D 1026

Search completed: April 6, 2005, 12:16:59
Job time : 48.6668 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 172.72 Seconds

(without alignments)
3459.921 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106

Sequence: 1 MELPFVTHLPFLVLTGLC.....GFFAHKKIPBEKREKLEQ 1167

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6106	100.0	1167	1 ITAG_HUMAN	O75578 homo sapien
2	2429	39.8	1189	1 ITAH_HUMAN	Q9ukx5 homo sapien
3	2418	39.6	1188	1 ITAH_MOUSE	P61622 mus musculus
4	2418	39.6	1188	2 Q7QCQ3	Q7qcq3 mus musculus
5	1894.5	31.0	1180	1 ITA1_RAT	P18614 rattus norv
6	1887	30.9	1151	1 ITA1_HUMAN	P56199 homo sapien
7	1853.5	30.4	1171	2 O42034	O42094 gallus gall
8	1756	28.8	1181	1 ITA2_HUMAN	P17301 homo sapien
9	1754	28.7	1170	1 ITA2_BOVIN	P53710 bos taurus
10	1744	28.6	1178	1 ITA2_MOUSE	Q62469 mus musculus
11	1743.5	28.6	1178	2 Q6P1C7	Q6p1c7 mus musculus
12	1487.5	24.4	823	2 Q8WY18	Q8wy18 homo sapien
13	1473	24.1	823	2 Q8CE84	Q8ce84 mus musculus
14	1273	20.8	288	2 Q8BM12	Q8bm12 mus musculus
15	1136	18.6	1161	1 ITAD_RAT	Q9gye7 rattus norv
16	1131.5	18.5	1162	1 ITAD_HUMAN	Q13349 homo sapien
17	1105	18.0	1170	1 ITAL_HUMAN	P20701 homo sapien
18	1097	18.0	1161	2 Q9WTV4	Q9wtv4 mus musculus
19	1097	18.0	1163	1 ITAL_MOUSE	P24063 mus musculus
20	1089.5	17.8	1160	2 Q9R200	Q9r200 mus musculus
21	1085.5	17.8	1188	2 Q6KAS4	Q6kas4 mus musculus
22	1079	17.7	1152	1 ITAM_HUMAN	P11215 homo sapien
23	1069	17.5	1169	1 ITAM_MOUSE	Q9qxh4 mus musculus
24	1054	17.3	1163	1 ITAX_HUMAN	P20702 homo sapien
25	1053.5	17.3	1165	1 ITAL_BOVIN	P61625 bos taurus
26	1026.5	16.8	1166	2 Q6TYB8	Q6tyb8 bos taurus
27	1022.5	16.7	1153	1 ITAM_MOUSE	P05555 mus musculus
28	1000.5	16.4	1151	2 Q9J130	Q9ji30 rattus norv
29	997	16.3	780	2 Q96271	Q96271 xenopus lae
30	984	16.1	1086	2 Q96HB1	Q96hb1 homo sapien
31	980.5	16.1	1196	2 Q98TF1	Q98tf1 cyprinus ca

32	953	15.6	1167	1 ITAE_MOUSE	Q60677 mus musculus
33	953	15.6	1167	2 O88340	O88340 rattus norv
34	947.5	15.5	1187	2 Q98TF0	Q98tf0 cyprinus ca
35	946	15.5	1167	2 O88341	O88341 rattus norv
36	932.5	15.3	1179	1 ITAE_HUMAN	P38570 homo sapien
37	891.5	14.6	1038	2 Q8BSQ1	Q8bsq1 mus musculus
38	883.5	14.5	920	2 Q28984	Q28984 sus scrofa
39	883	14.5	895	2 Q9WJF8	Q9wuf8 mus ep. itg
40	863	14.1	927	2 Q8HZV0	Q8hzv0 bos taurus
41	797	13.1	1332	2 Q9BPQ8	Q9bpq8 halocynthia
42	787.5	12.9	1160	2 Q8MKF4	Q8mkf4 felis silve
43	778	12.7	1036	2 Q91YD5	Q91yd5 mus musculus
44	738	12.1	1035	1 ITA9_HUMAN	Q13797 homo sapien
45	732.5	12.0	1054	2 Q9U6S1	Q9u6s1 strongyloce

ALIGNMENTS

RESULT 1

ID	ITAG_HUMAN	STANDARD;	PRT;	1167 AA.
AC	O75578; Q9UH28;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Integrin alpha-10 precursor.			
GN	Name=ITGA10;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Articular chondrocytes;			
RX	MEDLINE=98352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;			
RA	Camper L., Hellman U., Lundgren-Aakerlund E.,			
RT	"Isolation, cloning, and sequence analysis of the integrin subunit			
RT	alpha10, a beta1-associated collagen binding integrin expressed on			
RT	chondrocytes.";			
RL	J. Biol. Chem. 273:20383-20389(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Endothelial cells, and Heart;			
RX	MEDLINE=20169197; PubMed=10702680;			
RA	Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,			
RA	Wang S.-X., Langley R., Krissansen G.W.;			
RT	"The integrin alpha10 subunit: expression pattern, partial gene			
RT	structure, and chromosomal localization.";			
RL	Cytogenet. Cell Genet. 87:238-244(1999).			
CC	-!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.			
CC	-!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10			
CC	associates with beta-1.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: Widely expressed with highest expression in			
CC	muscle and heart. Found in articular cartilage.			
CC	-!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins			
CC	with I-domains do not undergo protease cleavage.			
CC	-!- SIMILARITY: Belongs to the integrin alpha chain family.			
CC	-!- SIMILARITY: Contains 7 FG-GAP repeats.			
CC	-!- SIMILARITY: Contains 1 VWFA domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF074015; AAC31952.1; -			
DR	EMBL; AF112345; AAF21944.1; -			
DR	EMBL; AF172723; AAF61638.1; -			

```

DR HSP; P18614; IMHP.
DR Genew; HGNC:6135; ITGA10.
DR MIM; 604042; -.
DR GO; GO:0008105; C:integrin complex; TAS.
DR GO; GO:0005518; F:collagen binding; TAS.
DR GO; GO:0007160; F:cell-matrix adhesion; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS02034; VWA; 1.
DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 1167 Integrin alpha-10.
FT DOMAIN 23 1122 Extracellular (Potential).
FT TRANSMEM 1123 1145 Potential.
FT DOMAIN 1146 1167 Cytoplasmic (Potential).
FT REPEAT 38 97 FG-GAP 1.
FT REPEAT 98 ? FG-GAP 2.
FT DOMAIN 167 350 VWA.
FT REPEAT 365 427 FG-GAP 3.
FT REPEAT 428 482 FG-GAP 4.
FT REPEAT 483 545 FG-GAP 5.
FT REPEAT 546 605 FG-GAP 6.
FT REPEAT 608 660 FG-GAP 7.
FT DOMAIN 1134 1140 Poly-Leu.
FT CA_BIND 494 502 Potential.
FT CA_BIND 558 566 Potential.
FT CA_BIND 620 628 Potential.
FT DISULFID 76 86 By similarity.
FT DISULFID 666 675 By similarity.
FT DISULFID 681 736 By similarity.
FT DISULFID 789 795 By similarity.
FT CARBOHYD 98 98 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 234 234 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 336 336 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 364 364 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 733 733 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 763 763 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 839 839 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 921 921 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1011 1011 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1018 1018 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1039 1039 N-linked (GlcNAc. .) (Potential).
FT CONFLICT 844 844 I -> L (in Ref. 2).
FT CONFLICT 909 909 G -> V (in Ref. 2).
FT CONFLICT 926 926 E -> D (in Ref. 2).
SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25C1AEAO CRC64;

Query Match
Best Local Similarity 100.0%; Score 6106; DB 1; Length 1167;
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELPFVTHLFLPLVLTGLCSFFNLDEHHPLFPQPEAEFGYSVLQHVGGQRWMLVGA 60
Db 1 MELPFVTHLFLPLVLTGLCSFFNLDEHHPLFPQPEAEFGYSVLQHVGGQRWMLVGA 60

Qy 61 PWDPSGDRRDVRCVPGGAHNAAPCAKXGHLGDYQLGNSSHPAVNMHLGMSLLTDDGG 120
Db 61 PWDPSGDRRDVRCVPGGAHNAAPCAKXGHLGDYQLGNSSHPAVNMHLGMSLLTDDGG 120

Qy 121 FMACAPLWSRACGSVFSSGICARVDASFQQSLAPTAQRCPTVMDVIVLDGNSIYP 180
Db 121 FMACAPLWSRACGSVFSSGICARVDASFQQSLAPTAQRCPTVMDVIVLDGNSIYP 180

Qy 191 WSEVQTFRLRLVGKLFIDPEQIQVGLVQGESPVHWSLGDPRFKKEVVRRAAKNLSREG 240

```

RESULT 2

ITAH HUMAN

ID ITAH HUMAN STANDARD; PRT; 1189 AA.

AC Q9UKX5; Q9UK01;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

Db 181 WSEVQTFRLRLVGKLFIDPEQIQVGLVQGESPVHWSLGDPRFKKEVVRRAAKNLSREG 240
Qy 241 RETKTAQAIWVACTEGFSQSHGGRPAARLLVVVTGESHGDBELPAALKACAGRVTRY 300
Db 241 RETKTAQAIWVACTEGFSQSHGGRPAARLLVVVTGESHGDBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHYLRQRDPSSFLREIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHYLRQRDPSSFLREIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVAGYDWGGSVLMLEGGHRLPPRMALEDEF 420
Db 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVAGYDWGGSVLMLEGGHRLPPRMALEDEF 420
Qy 421 PPALQNHAAVLYGYSVSMMLRGRRFLFSGAPFRHRGKVIAFQKKDGAVRVAQSLQGE 480
Db 421 PPALQNHAAVLYGYSVSMMLRGRRFLFSGAPFRHRGKVIAFQKKDGAVRVAQSLQGE 480
Qy 481 QIGSYFGSELCPDTRDGTVDVLLVAAMPFLGPQNKETGRVYVYLVGQQSLTLQGTQL 540
Db 481 QIGSYFGSELCPDTRDGTVDVLLVAAMPFLGPQNKETGRVYVYLVGQQSLTLQGTQL 540
Qy 541 PEPQDARFGFANGALPDLNQDGFADVAVGAPLEDGHQGALYLYHGTQSGVRPHPAORIA 600
Db 541 PEPQDARFGFANGALPDLNQDGFADVAVGAPLEDGHQGALYLYHGTQSGVRPHPAORIA 600
Qy 601 AASMPHALSYFGRSVDGRDLDDGDDVDVAVAGAAIILLSSRPIVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSYFGRSVDGRDLDDGDDVDVAVAGAAIILLSSRPIVHLTPSLEVTPOAIS 660
Qy 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAAFDGSGQ 720
Db 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALVTVPALDNTTKPGPVLNBSPTSIO 780
Db 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALVTVPALDNTTKPGPVLNBSPTSIO 780
Qy 781 KLVPPSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVGRRKVLVSTTLLENKENAYNT 840
Db 781 KLVPPSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVGRRKVLVSTTLLENKENAYNT 840
Qy 841 SLSIIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPVFTGAKVTFLLPEPFC 900
Db 841 SLSIIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPVFTGAKVTFLLPEPFC 900
Qy 901 SLLSQVFGKLTASSDSLENGTLQENTATQTSAYIQEPHLLFSSESTLHRYEVHPYGT 960
Db 901 SLLSQVFGKLTASSDSLENGTLQENTATQTSAYIQEPHLLFSSESTLHRYEVHPYGT 960
Qy 961 PVGPGPFKTLRVQNLGCVVSGLIISALLPAVAGGNYPFLSLSQVITNNASCIVQNL 1020
Db 961 PVGPGPFKTLRVQNLGCVVSGLIISALLPAVAGGNYPFLSLSQVITNNASCIVQNL 1020
Qy 1021 EPPGPPVHPELOHTNRLNGSNQCQVVRCHLQAGTSEVSGLLRLVHNEFPFRKFK 1080
Db 1021 EPPGPPVHPELOHTNRLNGSNQCQVVRCHLQAGTSEVSGLLRLVHNEFPFRKFK 1080
Qy 1081 SLTVVSTFELGTREGSVLQTEASRWSESLLEVVQVRPILISWILIGSVLGGILLALL 1140
Db 1081 SLTVVSTFELGTREGSVLQTEASRWSESLLEVVQVRPILISWILIGSVLGGILLALL 1140
Qy 1141 VFCLWKLGFPAHKKIPPEEKREKLEQ 1167
Db 1141 VFCLWKLGFPAHKKIPPEEKREKLEQ 1167

```

05-JUL-2004 (Rel. 44, Last annotation update)
 Integrin alpha-11 precursor.
 Name=ITG11;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Fetal heart, and Osteoblast;
 RX MEDLINE=99417678; PubMed=10486209; DOI=10.1006/geno.1999.5909;
 RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
 RA Wang S.-X., Morris C.M., Krissansen G.W.;
 RT "Cloning, sequence analysis, and chromosomal localization of the novel
 human integrin alpha11 subunit (ITG11).";
 RL Genomics 60:179-187(1999).
 [2]
 SEQUENCE FROM N.A.
 RC TISSUE=Fetal muscle, and Uterus;
 RX MEDLINE=99395147; PubMed=10464311; DOI=10.1074/jbc.274.36.25735;
 RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;
 RT "cDNA Cloning and Chromosomal Localization of Human alpha(11)
 Integrin. A collagen-binding, i domain-containing, beta(1)-associated
 integrin alpha-chain present in muscle tissues.";
 RL J. Biol. Chem. 274:25735-25742(1999).
 [3]
 SEQUENCE OF 954-1188 FROM N.A.
 RC TISSUE=Fibroblast;
 RA Andreu N., Estivill X., Escarceller M., Sumoy L.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen.
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11
 CC associates with beta-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: According Ref.2 highest levels in uterus and
 CC heart, intermediate levels in skeletal muscle and intermediate to
 CC low levels in pancreas, kidney and placenta. According to Ref.1
 CC also found in brain, colon, lung, small intestine, stomach,
 CC testis, salivary glands, thyroid glands and prostate. Very low
 CC levels in peripheral blood lymphocytes, fetal brain and fetal
 CC liver.
 CC -1- DEVELOPMENTAL STAGE: Strongly up-regulated in differentiating
 CC fetal muscle cells (in vitro).
 CC -1- DOMAIN: The integrin i-domain (insert) is a VWFA domain. Integrins
 CC with i-domains do not undergo protease cleavage.
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -1- SIMILARITY: Contains 1 VWFA domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF109681; AA01258.1; -
 CC EMBL; AF137378; AA051919.2; -
 CC EMBL; AL359064; CAB94392.1; -
 CC HSP; P18614; IMHP.
 CC Genew; HGNC:6136; ITG11.
 CC MIM; 604789; -
 CC GO; GO:0008305; C:integrin complex; TAS.
 CC GO; GO:0005518; F:collagen binding; TAS.
 CC GO; GO:0007160; P:cell-matrix adhesion; TAS.
 CC GO; GO:0007517; P:muscle development; TAS.
 CC InterPro; IPR00413; Integrin_alpha.
 CC InterPro; IPR002035; VWFA.
 CC Pfam; PF01839; FG-GAP; 3-
 CC Pfam; PF00092; VWFA; 1-
 CC PRINTS; PR01185; INTEGRINA.
 CC PRINTS; PR00453; VWFADOMAIN.

DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWFA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
 DR PROSITE; PS0234; VWFA; 1.
 KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium;
 KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.
 FT SIGNAL 1 22 Potential
 FT CHAIN 23 1189 Integrin alpha-11.
 FT DOMAIN 23 1142 Extracellular (Potential).
 FT TRANSMEM 1143 1165 Potential.
 FT DOMAIN 1166 1189 Cytoplasmic (Potential).
 FT REPEAT 38 94 FG-GAP 1.
 FT REPEAT 102 163 FG-GAP 2.
 FT DOMAIN 164 345 VWFA.
 FT REPEAT 359 420 FG-GAP 3.
 FT REPEAT 422 475 FG-GAP 4.
 FT REPEAT 477 537 FG-GAP 5.
 FT REPEAT 539 598 FG-GAP 6.
 FT REPEAT 601 653 FG-GAP 7.
 FT DOMAIN 1154 1162 Poly-Leu.
 FT DOMAIN 1174 1177 Poly-Arg.
 FT CA BIND 488 496 Potential.
 FT CA BIND 551 559 Potential.
 FT CA BIND 613 621 Potential.
 FT DISULFID 76 83 By similarity.
 FT DISULFID 121 139 Potential.
 FT DISULFID 129 159 Potential.
 FT DISULFID 659 668 By similarity.
 FT DISULFID 674 729 By similarity.
 FT DISULFID 781 787 By similarity.
 FT DISULFID 881 893 By similarity.
 FT CARBOHYD 82 82 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 95 95 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 291 291 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 331 331 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 358 358 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 449 449 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 462 462 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 528 528 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 642 642 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 694 694 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 857 857 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 894 894 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 973 973 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1032 1032 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1040 1040 N-linked (GlcNAc...) (Potential).
 FT VARIANT 433 433 V -> M.
 FT VARIANT 524 524 /FTid=VAR_009889.
 FT VARIANT 524 524 R -> L.
 FT VARIANT 972 972 /FTid=VAR_009890.
 FT VARIANT 972 972 L -> P.
 FT VARIANT 1003 1003 I -> M.
 FT VARIANT 1030 1030 /FTid=VAR_009892.
 FT VARIANT 1030 1030 Missing.
 FT VARIANT 1094 1094 /FTid=VAR_009893.
 FT VARIANT 1094 1094 L -> V.
 FT VARIANT 1094 1094 /FTid=VAR_009894.
 SQ SEQUENCE 1189 AA; 133609 MW; 60303C08AA4CD52 CRC64;
 Query Match 39.8%; Score 2429; DB 1; Length 1189;
 Best Local Similarity 43.1%; Pred. No. 3.5e-164;
 Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;
 QY 1 MELPVTHTLPLVFLTGLCSPEFLDEHHPRLPGPPAEAFGYSLQHVGGQRMVLVGA 60
 Db 1 MDLPGRGLVVAWALSLWFGFTDTTFNMDTRKFRVTPGSTRAPFGYTVQQHDSGNKWL VVGA 60
 QY 61 PWDGPGSDRRGDVYRCFVGGAHNAPCAKHGLGDYQLGNSSHPAVNMHGLMSLLETGDGG 120
 Db 61 PLEITNGYQKTGDVYKCFV---IHGNCCTKLNLGRVTLNSVSEKDNMRGLSLATNPKDNS 117
 QY 121 FMACAPLWSRACGSSVFSFSSGICARVDASFQPGQSLAPTAQRCPYMDVIVLDGNSIYP 180

Db 118 FLACPLWSEHCSSYTTGMCSSVNGSNFRFSKTVAPALQRCQTYMDIVVLGDSNIYP 177
Qy 181 WSEVQTLRLRLVGLKFLDPRQIOVLGVQYCESPVHWSLGDFTKKEVVRVAAKNLSREG 240
Db 178 WVEVQHFLILMLKFFLIGPOCIQGVVQYQVEDVHFEHLNDYRSVKDQVVEAAHSHIEQGG 237
Qy 241 RETKTAQIAIVACTEGFSQSHGGRPEARLLVVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 238 TETRTAFGIFSPARSEAFQK--GGRKGAKVMIVITDGESHSDPDLKVIQOSERNVTRY 295
Qy 301 GIAVLGHLRQRDPSSFLREIRTIASDPDRPFNVTDDBAALTIDIVDALGDRIPGLEGS 360
Db 296 AVAVLGYNRRGINPEFLNEIKYIASDPDKHFNVTDEAALKIDIVDALGDRIFSLEGT 355
Qy 361 HAENSEFGLMSQIGFSTRHLKDKGILFGWVGAVDMGSLVLEGGHRLPPRMALEDEF 420
Db 356 N-KNETSFGLEMSGTGFSSHVVEDVLGAVGAVDMGAVLKTSAGKVIPLRESYLKEF 414
Qy 421 PPALQNHAAVLGYSSVSMLLRGRRFLSLGAPRPHRGKVIATFOLKKGAVRVAQSLQGE 480
Db 415 PEELKXHGAVLYTIVTSVSSRQGRVYVAGAPRPHHTGKVLFTMHNRSLSLTHQAMRGQ 474
Qy 481 QIGSYFSGELCPLDTRDGTDLVLLVAAPMLGPNKQKTRGVYVVLGVQSSLLTLQOTLQ 540
Db 475 QIGSYFSGEITSVDIDGQVTDVLLVGAPMYFN--EGRERGKVVYVEL--QNRFFVYNGTLK 532
Qy 541 PEP--QDARGFGAMGALPDNLQDGFADVAVGAPLEDHOGALVLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNRFGSSIASVRDLNQSDVNDVVGAPLEDNHAGAIYIFHFGSILKTPQRI 592
Qy 600 AAASMPHALSYFGRSVDRGLDGDGLDVAVAGQAAILLSRPIVHLTPSLEVTPOAI 659
Db 593 TASELATGLQYFGCSINGQLDNEGLDLAVGALGNVILMSRPVVOINASHLFESKI 652
Qy 660 SVVORDCRRGQENAVCLTAALCFQVTSRTQGRMDHQPMFTASLDWTGARAAPFGSG 719
Db 653 NIFHRDCKRSGRDATCLAAFLCPTPIFLAPHFQTTVIGIRYNATMDERRVTPRAHLDEGG 712
Qy 720 QRLSPRLRLSVGNVTCQELHVEHLDTSDYLPRVALVTALDNTTKPGVILNKGSPSTI 779
Db 713 DRFTNRVAVLSSGQELCERINFHVLDTADYVKPVTFSVYSLEDPDH--GPMLDGQWPTTL 771
Qy 780 QKLVPPSKDCPDNECVTDVLQVNMDI-----RGRK-----APFVV 817
Db 772 RVSVFPWNGCNEDEHCVPLDVLARSDLPTAMEYQORVLRKPAQDCSAYTLSPDTTVFII 831
Qy 818 RGRKRVLVSTTLNRKENAYNTSLSIIFGRNLHLASLTQORSPPIKVECAAPS--AHAR 875
Db 832 ESTQRVAVEATLENRGENAYSTVNTSQSANLQFASLIQKSDSGSIECVNERRLQKQ 891
Qy 876 LCSVGHVPFQTKAVTFLLEFEFSCSLSQVFGKLTASSDSLERNCTLOENTAQTSAYI 935
Db 892 VCNVSYPPFRKAKAVAPRLDFFESKSIPLHLELTELAAAGSDSNRSDTKDNVAPLRFHL 951
Qy 936 QYEPHLLFSSESTLHRYEVPYGTLP--VGPGEFETKTLRVQLNGCVVSGLIISALLPA 993
Db 952 KYEADVLFTRSSLSHVEVLNSSLERYDGIQPPFCIFRIQNLGLFPIHGIMMKIPI 1011
Qy 994 VAHGNNVFLSQQVITN--NASC-IVQNTTEPPGPPVHPEELQHTNRLNGSNTOCQVVR 1049
Db 1012 ATRSGNRLKLRDLTDEVANTSCNIWGNSTEVPTPE--EDLRRAPQLNHSNDVVSIN 1070
Qy 1050 CHLGQLAGKEVSVGLLRVHNEFFRAKPSLTVVSTFELGTBEGSVLQLTASRWSES 1109
Db 1071 CNI--RLVPNGEINPHLLGNLWRLSKALKYKSMKIMVNAALQROFHPFFTFREEDPRQI 1129
Qy 1110 LLEVQTRPILISWILIGSLVGLGLLLALLVFCMLKGLGFFAHKKIPEEKKREKLE 1166
Db 1130 VFEISKQEDMQVPIWIIVGSTLGLLLALLVLLVLMKLGFFRSR-----RRREPLD 1182

ITAH MOUSE
ID ITAH MOUSE STANDARD; PRT; 1188 AA.
AC P61622;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-11 precursor.
GN Name=Itgall;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schnarch A., Schein J.E., Jones S.J.M., Warra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen (By
similarity).
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11
associates with beta-1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
with I-domains do not undergo protease cleavage.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; BC058716; AAH58716.1; --
CC MGD; MGI:2442114; Itgall.
CC PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS00234; VWFA; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
Repeat; Signal; Transmembrane.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 1188 Extracellular (Potential).
FT DOMAIN 23 1141 Extracellular (Potential).
FT TRANSMEM 1142 1164 Potential.
FT DOMAIN 1165 1188 Cytoplasmic (Potential).
FT REPEAT 38 94 FG-GAP 1.
FT REPEAT 102 163 FG-GAP 2.
FT REPEAT 164 345 VWFA.
FT DOMAIN 359 420 FG-GAP 3.
FT REPEAT 422 475 FG-GAP 4.
FT REPEAT 477 537 FG-GAP 5.
FT REPEAT 539 598 FG-GAP 6.
FT

FT	REPEAT	601	653	FG-GAP 7.	
FT	CA_BIND	488	496	Potential.	
FT	CA_BIND	551	559	Potential.	
FT	CA_BIND	613	621	Potential.	
FT	DISULFID	76	83	By similarity.	
FT	DISULFID	121	139	Potential.	
FT	DISULFID	129	159	Potential.	
FT	DISULFID	659	668	By similarity.	
FT	DISULFID	674	729	By similarity.	
FT	DISULFID	781	787	By similarity.	
FT	DISULFID	881	893	By similarity.	
FT	CARBOHYD	82	82	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	95	95	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	291	291	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	331	331	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	358	358	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	449	449	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	462	462	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	528	528	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	642	642	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	694	694	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	857	857	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	894	894	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	973	973	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	1031	1031	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	1039	1039	N-linked (GlcNAc).	(Potential).
FT	CARBOHYD	1059	1059	N-linked (GlcNAc).	(Potential).
SQ	SEQUENCE	1188 AA; 133011 MW; 0B0313C90D65422E CRC64;	39.6%; Score 2418; DB 1; Length 1188;		
Query Match					
Best Local Similarity 43.3%; Pred. No. 2.2e-163;					
Matches 512; Conservative 207; Mismatches 413; Indels 50; Gaps 15;					
Qy	18	GLCSFNLDDHRLPFGPPPEAEFGYSVLQVGGQRMVLPAPMDQPSGDRGDVYRCP	77		
Db	18	GFTDTFMDTRNPRVIAGPSAAPPYTVQQHDSIGKKWLTVGAPMETNGHKTGDVYKCP	77		
Qy	78	VGGNHNAPCAKHGLDQLGNSSHPAVNMHGLMSLLETDDGGGFMACAPLWSRACGSSVF	137		
Db	78	VTQGN---CTKLNLRVTLSNVSRKDNRLGLSLATNPKNDSFLACSLWSHECGSSYY	134		
Qy	138	SSGICARVDASFOFQSGSLPTAQRCPTVPMSEVQTFLLRLVGLFI	197		
Db	135	TTGWCSPANFRFSKTVAPALQRCQYMDIVIVLDGNSIYPVVEVQHFILNLIKPYI	194		
Qy	198	DPEIQVGLVQYGESPVHWSLGDPRTKKEEVRAAKNLSRREGRETKTAQAIMVACTEGF	257		
Db	195	PGQIQVGIQVYGEDA VHEFLNDYRSVKDVEEAASHIEQRGGTETRTAFGEIFARSEAF	254		
Qy	258	SQSHGRPEARLLVVVTDDGESHGDELPALKACEAGRVTRYGIAVLGHYLRQRDPSS	317		
Db	255	QK--GGRGKAKVMIVITDGESHSDPLEKVIQROSEKDNVTRYAVAVLGYNNRGINPET	312		
Qy	318	FLREIRTIASDPDEFNFVNTDEAALTDIVDALGDRIFGLEGSHAENESSFGLEMSQIGF	377		
Db	313	FLNEIKTIVADPDKHFNFVNTDEAALDKDIVDALGDRIFGLEGTN-KNETSFGLEMSQTGF	371		
Qy	378	STHRLKDGILFGMVGADWGSVLWLEGHRLFPFRMALEDEFPFALQNHAAVLYGYSVSS	437		
Db	372	SSHVEDGILLGAVGADWNGAVLKETISACKVIPHRESYLKEFFPEELKNHAAVLYGYVTS	431		
Qy	438	MLRGGRRLLFLSGAPRFRHKGKVIATFOLKKDGA VRAQSLQGEQIGSYFSGELCPDTR	497		
Db	432	VSSRQGRVTVAGAPRFRHGTGKVLFSMHNRSLSITHQALRGEQIGSYFSGEITSVDVND	491		
Qy	498	DGTTDVLVAPMFLPQNKQETGRVYVLYVQOQSLLTLQGLQPEPP-QDARFGFANGAL	556		
Db	492	DRVTDVLVAGPMYFS-EGREGRKVVYVNL-RQNRVYVNGTLKDSHSYQNAFSGCSIASV	549		
Qy	557	PDNLQDGFADVAVGAPLEDHQGALYLYHGTSQSVRPHPAQRIAAASMPHALSYFGRSVD	616		
Db	550	QDLNQDSYNDVVVAGPLEDHSHRGAIYIFHGFQTNILKKPMQRTIASELAQLOHFGCSIH	609		

Qy	617	GRLLDGGDLVDVAVGAQGAAILSSRPPIVHLTPSLEVTPQAISVVDORCRRRGQEAACL	676		
Db	610	QQLDNEGLVDLAVGALGNVAVLWARPVQVQINASLHFEFSKINIFHKDKCRNGRDACL	669		
Qy	677	TAALCFQVTSRTPGRWHDQFYMRFRTASLDEWTAGARAAFDGSGQRLSPRLRLSVGNVTC	736		
Db	670	AAFICFIPIFLAPHFQTATVIGIRYNATMDERRYMPRAHLDEGGDQFTNRAVLLSSGQSHC	729		
Qy	737	EQHLHFVLDTSYLRPVVALVTTFALDNTTKPGPVNLGSPSTSIQKLVFPFKDCCPDNECV	796		
Db	730	QRINFHVLDTADYVKPVAFVSVEYSLEDPDN-GPMLDNGWPTTLRVSPVFWNGCNEDEHCV	788		
Qy	797	TDLVQLVQNMDIRSK-----APFVVGGRKRVLSVTTLENRK	834		
Db	789	PDVLVDARSOLPTAMEYQCVLRPAQDCSSYTLSDFTTTFIIBSTRRRVAVEATLENRG	848		
Qy	835	ENAYNTSLSIIFSRNLHLASLTPORESPIKVECAAPS--AHARLCSVGHVPVFOGAKVTF	892		
Db	849	ENAYSAVINISQSENLOFASLIQKDDSDNSIECVNEERRLHKVCNVSYPPFRAKAKVAF	908		
Qy	893	LLPEFSCSLLSQVFGKLTASSDSLRNGTLQNTAQTSAIYQYEPHLLPSSSTLHRY	952		
Db	909	RLDFEFSKSVFLHHLQIHLGAGSDSHEQDSTADNTALLRFLHKYEADVLFTRSSLSHF	968		
Qy	953	EVHPYGTLPV--GPGPEFKTTLRVQNLGCVVSGLIISALLPAVAHGSNYFLSLSQVITN	1010		
Db	969	EVKANSLESYDGIIGPPFNCVFNQNLGFFPIHGVMKMTVPIATRGNLLMLURDPTD	1028		
Qy	1011	--NASC-IVONLTPEPPGVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067		
Db	1029	QGNWSCNINWNSYEYRSTPTE-EDLSHAPQNHNSNDVSVIICNL-RLASQETSFVYLG	1086		
Qy	1068	LVNEFFPRRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLLEVQVTRPILSLMILI	1127		
Db	1087	NLMTLSLKALKYRSKITVNAALQRPSPFIREDPSRQVTEISKQEDWQVPIWIIIV	1146		
Qy	1128	GSVLGGLLALLAVFLCWLKLGFFAHKK-----IPEEEK	1160		
Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGLGPIPKBLK	1188		
RESULT 4					
ID	Q7TQC3	PRELIMINARY;	PRT;	1188 AA.	
AC	Q7TQC3;				
DT	01-OCT-2003	(TrEMBLrel. 25, Created)			
DT	01-OCT-2003	(TrEMBLrel. 25, Last sequence update)			
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)			
DE	All integrin.				
GN	Name=Itgail;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	PubMed=1518724; DOI=10.1016/j.ydbio.2004.03.006;				
RA	Popova S.N., Rodriguez-Sanchez B., Liden A., Betscholtz C.,				
RA	van den Bos T., Gullberg D.;				
RT	"The mesenchymal alpha1beta1 integrin attenuates PDGF-BB-stimulated				
RL	chemotaxis of embryonic fibroblasts on collagens.";				
RL	Dev. Biol. 270:427-442 (2004).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Johaneson M., Popova S.N.;				
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.				
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).				
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.				
DR	EMBL; AY124460; AAM62130.1; -				
DR	HSSP; P18614; ICK4.				
DR	GO; GO:0006929; P:substrate-bound cell migration; IMP.				
DR	InterPro; IPR000413; Integrin_alpha.				
DR	InterPro; IPR002035; VWF_A.				

```
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; VWA; 1.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; VWA; 1.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 1188 AA; 133067 MW; 69B2BF20D75E4271 CRC64;

Query Match          39.68; Score 2418; DB 2; Length 1188;
Best Local Similarity 43.38; Pred. No. 2.2e-163;
Matches 512; Conservative 207; Mismatches 413; Indels 50; Gaps 15;

QY 18 GLCSFNLDDHRLPFGPPEAFEGYSVLQVHGQGRWMLVGPWDPGSDRGRDGVYRCP 77
DB 18 GFTDTFNMDRNPRVIAAGPSAAFFGYTVQQHDISGKKWLVGAPMETNGHKQTKDGVYKCP 77
QY 78 VGAHNAAPCAKHGLDYLQNSHPAVNMHLGHSLLSTDDGGDFMACAPLWSRACGSVF 137
DB 78 VTQGN---CTKJNLGRVYTLGNVSRKDNMRGLGLSLATNPKNDSNPLACPLWSHECGSSYY 134
QY 138 SSGICARVDASFOPOGSLAPTAQRCPTMDVTVLDGNSIYPWSEVQTFRLRLRVGLFI 197
DB 135 TTGWCVRNENFRFSKIVAPALQRCQTYMDIVVLDGNSIYPWSEVQHFLLINLKFFYI 194
QY 198 DPEIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSRREGRETXTAQAIMVACTEGF 257
DB 195 GPGQIQVGIQVQGEDA VHEPLNDYRSVKDVEAASHIEQGGTETRTAFGIEFARSEAF 254
QY 258 SQSHGGRPEARLLVVTDESHDGELPAALKACEAGRTRYGIAVLGHYLRQRDPSS 317
DB 255 QK--GGRKGAKVMIVITDESHDSPLEKVIQSEKDNVTRYAVAVLYGYNRRGINPET 312
QY 318 FLREIRTIASDPERPFPNTVDEAALTDIVDALGDRIFGLEGSHAENESSFGLEMSOIGF 377
DB 313 FLNEIKIYASDDPKHFNFTVDEAALDXIDVALGDRIFSLGNT-KNETSGLEMSGTGF 371
QY 378 STHRLKDGILFGMVGAYDWSGLVLEGGHRLFPFRNALDEFPFALQNAHAAYLGYSVSS 437
DB 372 SSWVEDGILLGAVGYDMNGAVLKETSAGKVIPIHRESYLKERPEELKNHVAIYGYTVTS 431
QY 438 MLRGGRRLFLSGAPRRHKGKVIAPQLKDGAVRVAQSIQGGQIGSYFGSELCPLDTR 497
DB 432 VSSRQGRVTVAGAPRPHNTGKVLFSMHNRSITIHQALRGEQIGSYFGSEITSDVDND 491
QY 498 DGTDTDLVLAAPFLGQNKTRGVYVLYVQGSLLTLQGTLOPEPP-QDARGFAMGAL 556
DB 492 DRTVDVLLVGAAPHYS-EGRGGKVIYVNL-RQNRVYVNTGLKDSHYQNRGFCISV 549
QY 557 PDLNQGFADVAVGAPLEDHQGALYLYHGTQSGVRPHPAQRIAAASMPHALSYFGRSVD 616
DB 550 QDLNQDSYNDVVGAPLEDHSGAIYIFHGFQTNILKPKMQRITASELAPQLQHFGCSIH 609
QY 617 GRLLDQDGLVDVAVGAGQAAILLSRPVHLTPSLVTPQALISVVORDCRRRQBAVCL 676
DB 610 QGLDLNEDGVLAVGALGNAVLWARPVYQINASLHFEPKSKINIPHKDCKRGRDATCL 669
QY 677 TAALCFQVTSKTPGRWDHQFMRPTASLDSEWAGARAAAFDGGSGORLSPRLRLSVGNVTC 736
DB 670 AAFCLFPIPLAHPQTATGIRYNALMDERRYPRAHLDEGGDQFTNRAVLSSGQEH 729
QY 737 EQLHFHVLDTSDYLRPVALVTFTALDNTKPGPVINEGSPSTSIQKLVFSSKDCGPDNECV 796
DB 730 QRINFHVLDTADYVKKVAFSVSEYSLDDPN-GPMLDNGWPTTLRVSVFPWNGCNEDEHCV 788
QY 797 TDLVLQVNMIRGRK-----APFVVGGRKRVLVSTLTENRK 834
DB 789 PDLVLDARSDLPTAMEYCORVLGRPAQDCSSYTLSPDTFTVFIETSTRRRVAVENTLENRG 848
QY 835 ENAYNTSLSTIFSNNLHSLTTPORESPIKVECAAPS--AHARLCVSHGVFOTGAKVTF 892
DB 849 ENAYSAVLNLSQSENLOFASLIQKDDSDNSIECVNEERLHKVCNVSYFFFRKAKAVAF 908
QY 893 LLEFEFSCSLLSQVFGKLTASSDLSLRNGTLQNTAQTSAITYEPHLLFSSSESTLHRY 952
```

```

909 RLDPEFSKSVFLHHLQIHLGAGSDSEHQDSTADDNTALLRFLKYEADVLFTRSSLSHP 968
953 EVHPYGTLPV--GPGPEPKTTLAVQNLGVYVSGLLISALLPAVAGGNYFLSLSQVITN 1010
969 EVKANSLSYDYGIPGPPNCFVKVQNLGFFPIHGVNMKITVPTATRGNNLLMLRDPFTD 1028
1011 --NASC-IVONLTPPPVHPPELOHTNRLNGSNTOCCVVRCHIGQLAKGTVSVGLLR 1067
1029 QGNTSNINWNSTEYSTTPE-EDLSHAPQRNHSNDVSIICNL-RLAPSQTSFYLVG 1086
1068 LVHNEFRRAKFKSLTVSTFELGTSEGSVLQTEASRMSSESLLEVQVTRPILISLWILI 1127
1087 NLWLSLKALKYSLKITVNAALQRFHPFPFREEDPSQVTFEISKQEDMQVPIIIV 1146
1128 GSVLGGILLALLVFLWLKGLGFFAHKK-----IPEEK 1160
1147 GSTLGGILLALLVFLWLKGLGFFAKRREPGLGPIPKELK 1188

RESULT 5
ID _ITALI_RAT STANDARD; PRT; 1180 AA.
AC P18614;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Integrin alpha-1 precursor (Laminin and collagen receptor) (VLA-1)
DE (CD49a).
GN Name=Itgal;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90338125; PubMed=2380249; DOI=10.1083/jcb.111.2.709;
RA Ignatius M.J., Large T.H., Houde M., Tawil J.W., Barton A., Esch F.,
RA Carbonetto S., Reichardt L.F.;
RT "Molecular cloning of the rat integrin alpha 1-subunit: a receptor for
RT laminin and collagen."
RL J. Cell Biol. 111:709-720(1990).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.
RX MEDLINE=99313197; PubMed=10386626; DOI=10.1016/S0014-5793(99)00666-3;
RA Nolte M., Pepinsky R.B., Vanyaminov S.Y., Kotliansky V.,
RA Gotwals P.J., Karpusas M.;
RT "Crystal structure of the alpha1beta1 integrin I-domain: insights into
RT integrin I-domain function."
RL FEBS Lett. 452:379-385(1999).
CC -1- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and
CC collagen. It recognizes the proline-hydroxylated sequence G-P-P-G-
CC E-R in collagen.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1
CC associates with beta-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VWFA domain.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52140; CAA36384.1; -.
CC PIR; A35854; A35854.
CC PDB; 1CK4; X-ray; A/B=167-364.
DR
DR
```

DR PDB; 1MHP; X-ray; A/B=169-360.
DR RGD; 2923; Icgai.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PRO1185; INTEGRIN.
DR PRINTS; PRO453; VWFADOMAIN.
DR SMART; SM00327; VWA; 1.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWF; 1.
KW 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;
KW Magnesium; Receptor; Repeat; Signal; Transmembrane.
FT SIGNAL 1 28
FT CHAIN 29 1180 Integrin alpha-1.
FT DOMAIN 29 1142 Extracellular (Potential).
FT TRANSMEM 1143 1165 Potential.
FT DOMAIN 1166 1180 Cytoplasmic (Potential).
FT REPEAT 104 103 FG-GAP 1.
FT REPEAT 144 ? FG-GAP 2.
FT REPEAT 175 388 VWF.
FT DOMAIN 175 388 FG-GAP 3.
FT REPEAT 377 432 FG-GAP 4.
FT REPEAT 433 484 FG-GAP 5.
FT REPEAT 485 565 FG-GAP 6.
FT REPEAT 567 626 FG-GAP 7.
FT REPEAT 629 681 Potential.
FT CA_BIND 497 505 Potential.
FT CA_BIND 579 587 Potential.
FT CA_BIND 641 649 Potential.
FT SITE 1168 1172 GFFKR motif.
FT DISULFID 82 92 By similarity.
FT DISULFID 687 696 By similarity.
FT DISULFID 702 755 By similarity.
FT DISULFID 807 813 By similarity.
FT DISULFID 877 885 By similarity.
FT DISULFID 1029 1062 By similarity.
FT DISULFID 1066 1073 By similarity.
FT CARBOHYD 100 100 N-linked (GlcNAc. .)
FT CARBOHYD 105 105 N-linked (GlcNAc. .)
FT CARBOHYD 112 112 N-linked (GlcNAc. .)
FT CARBOHYD 217 217 N-linked (GlcNAc. .)
FT CARBOHYD 317 317 N-linked (GlcNAc. .)
FT CARBOHYD 341 341 N-linked (GlcNAc. .)
FT CARBOHYD 402 402 N-linked (GlcNAc. .)
FT CARBOHYD 418 418 N-linked (GlcNAc. .)
FT CARBOHYD 459 459 N-linked (GlcNAc. .)
FT CARBOHYD 531 531 N-linked (GlcNAc. .)
FT CARBOHYD 531 531 N-linked (GlcNAc. .)
FT CARBOHYD 698 698 N-linked (GlcNAc. .)
FT CARBOHYD 747 747 N-linked (GlcNAc. .)
FT CARBOHYD 779 779 N-linked (GlcNAc. .)
FT CARBOHYD 820 820 N-linked (GlcNAc. .)
FT CARBOHYD 839 839 N-linked (GlcNAc. .)
FT CARBOHYD 882 882 N-linked (GlcNAc. .)
FT CARBOHYD 907 907 N-linked (GlcNAc. .)
FT CARBOHYD 938 938 N-linked (GlcNAc. .)
FT CARBOHYD 965 965 N-linked (GlcNAc. .)
FT CARBOHYD 973 973 N-linked (GlcNAc. .)
FT CARBOHYD 1007 1007 N-linked (GlcNAc. .)
FT CARBOHYD 1084 1084 N-linked (GlcNAc. .)
FT CARBOHYD 1103 1103 N-linked (GlcNAc. .)
FT CARBOHYD 1114 1114 N-linked (GlcNAc. .)
FT STRAND 172 178 N-linked (GlcNAc. .)
FT HELIX 186 197
FT TURN 198 199
FT TURN 204 205
FT STRAND 208 214
FT STRAND 218 222
FT TURN 224 226
FT HELIX 230 238
FT TURN 239 239
FT HELIX 250 259

FT TURN 260 261
FT HELIX 264 266
FT TURN 267 267
FT TURN 270 271
FT STRAND 274 280
FT TURN 285 286
FT HELIX 287 289
FT HELIX 290 299
FT TURN 300 301
FT STRAND 302 309
FT HELIX 311 315
FT TURN 316 317
FT HELIX 321 328
FT TURN 329 330
FT HELIX 335 338
FT STRAND 339 342
FT HELIX 346 350
FT HELIX 352 359
SQ SEQUENCE 1180 AA; 130808 MW; 8E5DA2BE02362EE4 CRC64;
Query Match 31.0%; Score 1894.5; DB 1; Length 1180;
Best Local Similarity 36.5%; Pred. No. 5.6e-126;
Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;
QY 13 LVFLTGLCSPLNLDHHPRLPPGPEAEFGYSVLQHVGGQRWMLVGA PWDPSPGDRGD 72
DB 19 LTVLGFCVSNFVDVKNMSFSGVEVDMFGYTVQOQYENEEGKWLIGSPLVGVGPKARTGD 78
QY 73 VYRCVGGAHNAPCAKGLHGDYQLGNSSHPAV-----NMHLGMSLLETGDCGCFMACAPL 127
DB 79 VYKCPVGRERAMPVCVLDLP-----VNTSIPNVTEIKENMTFG-STLVNPNNGGLACGPL 133
QY 128 WSRACSSVSFSGICARVDASFPQGSGLAPTAQRCPTVMVIVLDGNSNTPWSEVQTF 187
DB 134 YAYRCGHLHTVTGICSDVSPFQVNSFAP-VQECSTQLDIVIVLDGNSNTPWSEVIAF 192
QY 188 LRLVGLKFLDPQIQVLQVQGESPVHWSLGDPTKEEVVRAAKNLSRREGRETAKQ 247
DB 193 LNDLLKRMIDGPKQTQVGIQVQGENVTHEFNLNKSYSTEELVVAANKIGRQGGIQTMTAL 252
QY 248 AIMVACTGFSQSHGGRPEARLLVVYTDGSHDGEELPAALKACEAGRVTRYGLAVLGH 307
DB 253 GIDTARKEATEARGARRGVKVMVIVTDGSHDNRYLKVQIQDCEDENITQFSAIALGH 312
QY 308 YLRRQRDPSPFLREIRTIASDPDERFFNVYDDEAALTDIVDALGDRIFGLGSHAESNESS 367
DB 313 YNRGNLSTEFVBEIKSIASEPTEKHFNFVSDDELAVTIVKALGERIFALEATADQSAAS 372
QY 368 FGLEMSQIGFSTHRLKDGILFGMVGAYDWGSSVLWLEGGHRLFPFPRMALEDEFPPALQNH 427
DB 373 FEMEMSQTFSAHYSQDWMVLMGAYDWMGTVMVMQANQWVIPHNTTTFOTE--PAKWE 430
QY 428 --AAYLCYSVSSMLLRGGRRLFLSGAPFRHRGVIAFQOLKKGAVRVAQSLQGEQIGSY 485
DB 431 PLASLYGTVNSATIPGD-VLYIAGQPRYNHTGVQVIYKM-EDGNINILQTLGGEQIGSY 488
QY 486 FGSELCLPDLTDGTTDLVLAAPMFLGPQNKETGRVYVYLVGQOSSLTLTGTLQP----- 541
DB 489 FGSVLTITIDDKDSYTDLLLVGAPMYNGTEKEEGKVYVAV-NQTRFEYQMSLEPIRQT 547
QY 542 -----EPPQDARFGFAMGALPDLNODGFADVAVGAPLEDHGGQALYLYHG 586
DB 548 CCSLKDNSCTKENKNEPCGARFGTAIAAVKDLNVDGNDVVI GAPLEDDHAGAVIYHG 607
QY 587 TQSGVRPHPAQRTAAASMPHALSYFGRSVDRGLDDGDLVDVAVGAOAGAILLSRPV 646
DB 608 SGKTIREAYAQIPSGGGDKTLKFFGOSIHGEMDLNGDLTDVTIGGLGGHALFWARDVA 667
QY 647 HLTPSLEVTPQAISVQVQDCRRRGOEAVCLTAALCFQVTSRTPGRWHDQFTRFASLIDE 706
DB 668 VVKVTMNFEPKNVNIQKKNCRVEGKETVCINATWCFHVKLSKEDSIYEADLQYRVTLDS 727
QY 707 WTAGARAAFDGSGQRRLSPRLRLSLVGNVTCEQLHFHVLDTSDYLRPVALVTFFALDNTTK 766

```
Db 728 LRQISRFPGTQERKIQ--NITVRESECIHRSFYMLDKHDFQDSVRVTLDLFD---NLTD 782
Qy 767 P--GPVLNEGSPSIOKLFPKDCGPDNCDVLDVQVNDIRGSRKAPFVVRGGRKV 824
Db 783 P--GPVLDDALPNSVHIEHPADCKGKERCISDLTLNVST-----TEKSLIIKVSQHDKF 838
Qy 825 LVSTTLNKRKENAYNTSLSIIFRNHLASLTP-QRESPIKVECAAPSAHARLCSVGHVP 883
Db 839 NVSLTVNKGDSAYNTRTVQHSPLNLFSGIEBIQKDS-----CESNQNITCRVGYPF 891
Qy 884 FQGAQVTFLEPFCSSLLSQVFKLTASSLSLRNGTLQNTAGTSAYIOVEPHILF 943
Db 892 LRAGETVTFKIIIFQNTSHLSNAIHLSATSDSEEPLESLNDNEVNSISFVRYEVLQF 951
Qy 944 SSSTLHRYEHPVGTLP-----VGPGRPKTLRVQNLGCVVSGSLIISALLP-AVAH 996
Db 952 YSSASEHHSIVANETIPEFINSTEDIGNEINVFYIRKRGHPMPELQLSISFPNUTAD 1011
Qy 997 GGNVFLSLQVITNNAQVQNLTEPPG-----PPVHPEELQHTNRLNGSNTQCQVVR 1050
Db 1012 GYPVLYPIGSSNDVNCPRSLDPFGINSKQKWTISKSEVLKRGTIQDCSSTCGVATI 1071
Qy 1051 HLQQLANG-TEVSVGLLRLVHNEFFRPAKPSLTIVTSTFELGTGEGSVLQLTASRWSES 1109
Db 1072 TCSLLPSDLGQVNVSL--LLWKPTFIRAFPSLSMLTLRGELKSENS-LTSSSNRKREL 1128
Qy 1110 LLEVVPQ-RPILSLTLGSLVGLLALLVFLWKLGLGFAHKKTPSEKREE 1163
Db 1129 AIQSKDGLGRVPLWILLISAFAPLMLLLMLLALWKIGFF---KRPLKKOMEK 1180

RESULT 6
ITAL_HUMAN
ID _ITAL_HUMAN STANDARD; PRT; 1151 AA.
AC P56199;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
GN Name:ITGAL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=3115124; PubMed=8428973;
RA Briesewitz R., Epstein M.R., Marcantonio E.E.;
RT "Expression of native and truncated forms of the human integrin alpha
RL J. Biol. Chem. 268:2989-2996(1993).
CC -1- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and
CC collagen. It recognizes the proline-hydroxylated sequence G-F-P-G-
CC E-R in collagen.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1
CC associates with beta-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VWFA domain.
CC -1- DATABASE: NAME=PROW; NCBI-CD guide CD49a entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49a.htm".
DR PIR; A45226; A45226.
DR PDB; 1QCS; X-ray; A/B=140-331.
DR PDB; 1QCY; X-ray; A=141-333.
DR Genew; HGNC:6134; ITGAL.
DR MIM; 192968; -.
DR GO; GO:0008105; C:integrin complex; TAS.
DR GO; GO:0005518; F:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; NAS.
```

```
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWFA; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR PRINTS; PRO0453; VWFA; 1.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
DR 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;
DR Magnesium; Receptor; Repeat; Transmembrane.
KW Extracellular (Potential).
FT DOMAIN 1 1113
FT TRANSMEM 1114 1136
FT REPEAT 1137 1151
FT REPEAT 16 75
FT REPEAT 76 ?
FT DOMAIN 147 360
FT REPEAT 349 404
FT REPEAT 405 457
FT REPEAT 459 520
FT REPEAT 540 599
FT REPEAT 602 654
FT CA BIND 470 478
FT CA BIND 552 560
FT CA BIND 614 622
FT SITE 1139 1142
FT DISULFID 54 64
FT DISULFID 660 669
FT DISULFID 775 728
FT DISULFID 780 786
FT DISULFID 850 858
FT DISULFID 1002 1034
FT DISULFID 1037 1044
FT CARBOHYD 46 46
FT CARBOHYD 72 72
FT CARBOHYD 77 77
FT CARBOHYD 84 84
FT CARBOHYD 189 189
FT CARBOHYD 289 289
FT CARBOHYD 313 313
FT CARBOHYD 374 374
FT CARBOHYD 390 390
FT CARBOHYD 432 432
FT CARBOHYD 504 504
FT CARBOHYD 671 671
FT CARBOHYD 720 720
FT CARBOHYD 752 752
FT CARBOHYD 812 812
FT CARBOHYD 855 855
FT CARBOHYD 880 880
FT CARBOHYD 911 911
FT CARBOHYD 938 938
FT CARBOHYD 946 946
FT CARBOHYD 980 980
FT CARBOHYD 1045 1045
FT CARBOHYD 1055 1055
FT CARBOHYD 1074 1074
FT CARBOHYD 1085 1085
FT STRAND 143 150
FT TURN 153 154
FT TURN 158 158
FT TURN 159 169
FT TURN 170 171
FT TURN 174 174
FT TURN 176 177
FT TURN 179 186
FT STRAND 190 194
FT TURN 196 197
FT HELIX 202 211
FT HELIX 222 231
```

```

FT TURN 232 234
FT HELIX 236 238
FT TURN 239 239
FT TURN 242 243
FT STRAND 245 252
FT TURN 257 258
FT HELIX 259 261
FT HELIX 262 271
FT TURN 272 273
FT STRAND 274 281
FT HELIX 283 287
FT TURN 288 289
FT HELIX 293 302
FT HELIX 307 310
FT STRAND 311 314
FT HELIX 317 323
FT HELIX 324 330
SQ SEQUENCE 1151 AA; 127837 MW; 6B3F3C1AABF52808 CRC64;

Query Match 30.98; Score 1887; DB 1; Length 1151;
Best Local Similarity 36.34; Pred. No. 1.9e-125;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;

QY 23 FNLDEHHPLRFPGPPEAEFGVSLQHVGGGQRMMLVGAHPWDGPGDRRGDVYRCPVCGAH 82
DB 1 FNVVKNMTFSGPVEDMFGYVQYENEBGKVLGSLVPGQPKNRTGDKYKCPVRGE 60
QY 83 NAPCAKHLGDYQLGNSHPAV-----NMHLGMSLLETDGDDGPFMACAPLWSRACGSSVF 137
DB 61 SLPCVKLDLP-----VNTSIPNVTVEKNMTFG-STLVTPNPGGFLACGPLYAVRCGHLHY 115
QY 138 SSGICARVDASFQPGQSLATQAQCPTVMYDVIVLDGNSIYPWSEVQTLRLVLGKLF 197
DB 116 TTGICDSVSTFTQVNVSIAP-VQECSTQLDIVIVLDGNSIYPWSDVTAFLNDLLKMDI 174
QY 198 DPEIQVGLVOYGESPVHWSLGDFTKEEVVRAAKNLSRREGRETKTAQAIMVACTEGF 257
DB 175 GPKQTQVGIVOYGENVTHEFNLNKYSSTEEVLVAKKIVQGGQRTWTALGTDTRKEAP 234
QY 258 SQSHGGRPEARLAVVVTGDGSHDGEELPAALKAACEAGRVTTRYIAVLGHYLRQRDPSS 317
DB 235 TEAGARRGVKVMVITVDGESHNDHRLKKVIQDCEDENIQRFSAIILGSYNRGNLSTEK 294
QY 318 FLBRIITADPDRFFNTVNTDEAALTDIVDALGDRIFGLEGSHAENESSFLEMSQIGF 377
DB 295 FVEEIKSIASEPTEKHFNVSDALAVTIKTLGERIFALEATADQSAASFEMEMSQTF 354
QY 378 STHRLKDGILFGMVGADWGSVLWLEGGHRLFPPRMALDEPPPALQNHAAVLGYSVSS 437
DB 355 SAHYSQWVLMGAVGADWGTVMVQKASQIIIPRNTTFNVESTKKNPEPLASYLYTVNS 414
QY 438 MLLRGGRRLLFSGAPFRHRGKVIATFQKKDGAVRVAQSLQGEQIGSYFGSELCPBLDTR 497
DB 415 ATASSGDVLVITAGQPRYNTHTGVIIYRM-EDGNIKILQTLGSEQIGSYFGSILLTTDIDK 473
QY 498 DGTDDLVLVAAPMLPQNKETGRVYVYLVCQQSLLTLQGLQP----- 541
DB 474 DSNITDILLVAPMTGMTEKEEQGVVYVAL-NOTRFEYQMSLEPIKOTCCSSROHNSCTT 532
QY 542 ---EPPDARFGFAMGALPDNLQDGFADVAVGAPLEDHOGALVYHGTQSGVPRPAPQR 598
DB 533 ENKNEPCGAPFGTAIAVKDLNDFNDIVIGAPLEDHGGAVIYHGSQTKIRKEVAQR 592
QY 599 IAAASMPHALSYFGSRVDRGLDGLDVLVDVAVGAQGAAILSSRPVHLTPPSLEVTPQA 658
DB 593 IPSGGDGKTLKFFQOSIHGENDLNGDGLTDVTIGLGGALFWSRDVAVKVTVNFPENK 652
QY 659 ISVVQRDCRRRGQAVCLTALACFQVTSRTPRGWDHOFYMRFTASLEWTAGAATAFGS 718
DB 653 VNIQKXCHMEGKETVCINATVCPCEVKLSKSDIETIYEADLQYRVTLDSLRQISRSFSGT 712
QY 719 GQRLSPRLRLSVGNVTCQLHFVHVLQTSVLRVVALTVFALONTYKPG- GPVLNBSGP 776
DB 719 GQRLSPRLRLSVGNVTCQLHFVHVLQTSVLRVVALTVFALONTYKPG- GPVLNBSGP 776

```

```

713 QERKQOR--NITVRKSECTKHSMFLDKHDFQDSVRITLDF---NLTPDENGPVLDSDSLP 767
777 TSIOQLVPPSKDQDNECVTDVLQVNMDIRSGRKAPFVVRGGRKVLSTTLENKEN 836
768 NSVHEYIPFAKDCGKKEKICISLHLV-----ATTEKDLLIVRSQNDKFNVSITVKNKDS 823
837 AYNTSLSIIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPVQTCAGKVTFLLEF 896
824 AYNTRTIVHSPNLVFSGI-----EAIQKDSG--ESNHNITCKVGYFPLRGEWTFKLLP 877
897 EFSCSLLSQVFGKLTASSDSLERNGTLQENTAQTSAYIOVEPHLLFSSESTLHRYEVHP 956
878 QFNTSYLMENVTVIYLSATSDSEBPPETLSDNVNVISIPVKEVGLQFVSSASEVHISIAA 937
957 YGTLDP-----VGPGEFKTLRVQNLGCVVSGLIISALLPAVAHGN-----YFLSLSQV 1007
938 NETVPEVINSTEDIGNEINIFYLIRKSGSPMPPELKLISFPNMTSNGYPVLYPTGLSS- 996
1008 ITNASCIVQNLTEP-----PGPPVHPEELOHTNRLNGSNTOCQVVRCHLQGLAKGTEV 1061
997 -SENANCRPHIFEDPPPSINGSKQMTTSTDLHKRGTILDCTCKEATTCNLTSS-SDISQV 1054
1062 SVGLLRVLVHNEFFRRRAKFKSLTVVSTFELGTGEEGVLQLTEASRWSLLEVVQT-RRPIL 1120
1055 NVSL--ILWKPTFIKSYFSSNLNLTIRGELRSENAS-LVLSNNQKRELAIOISKDGLPGR 1111
1121 ISWLILGSLVGLLLALLVFLCWLKGLPFAHKKIPPEEKREE 1163
1112 VPLVILLSAFAGLLLLMLLILALWKIGTF---KRPLKKRMEK 1151

RESULT 7
O42094 ID O42094 PRELIMINARY; PRT; 1171 AA.
AC O42094;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Alphas integrin.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gizzard;
RX MEDLINE=97476270; PubMed=9334246; DOI=10.1074/jbc.272.42.26643;
RA Obata H., Hayaishi K., Nishida W., Momiyama T., Uchida A., Ochi T.,
RA Sobue K.;
RT "Smooth muscle cell phenotype-dependent transcriptional regulation of
RT the alphas integrin gene.";
RL J. Biol. Chem. 272:26643-26651(1997).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AB000470; BAA23160.1; -.
DR PIR; A55348; A55348.
DR HESP; P18614; IMHP.
DR GO; GO:0008305; C: integrin complex; IEA.
DR GO; GO:0007160; P: cell-matrix adhesion; IEA.
DR GO; GO:0007229; P: integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00357; FG-GAP; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00234; VWF_A; 1.
KW Cell adhesion; Integrin; Transmembrane.

```

8Q	SEQUENCE	1171 AA; 130228 MW; B505A4D65F09736E CRC64;	
	Query Match	30.4%; Score 1853.5; DB 2; Length 1171;	
	Best Local Similarity	34.7%; Pred. No. 4.7e-123;	
	Matches	416; Conservative 236; Mismatches 464; Indels 83; Gaps 21;	
QY	13	LVFLTGLCSFNLDEHHPRLFPGPEAEFGYSVLQHVGGQRMWLVGAPWDGFGSGDRGD 72	
DB	8	LTALFHLIGASNDVDKNAITFSGPLEDMFGVTVOYENEEGKWVLIGSLVPGQPEKRTGD 67	
QY	73	VYRCPUGANNAACAKHLDG-VQLGNSHPAVNMHGLSLELTDGDDGGMACAPLWSRA 131	
DB	68	VYKCPVGRDSQPCIKNLDPATSVPNVMEVKENMTLGTTLV-TNPKGFGFLACQPLIYAYK 126	
QY	132	CGSSVRSFGICARVDASGFOQGLAPTAQRCPVMDVVIVLDGNSIYPWSEVQTPRLRL 191	
DB	127	CGRLHYTTGVCNSVSTFETVKAAPSVEQCKQLDIVIVLDGNSIYPWSEVTAFLNSL 186	
QY	192	VGKLFIDPEIQVGLVQYGSVPVHWSLGDPRTKBEVYVRAAKNLSRREGRETTKTAQIMV 251	
DB	187	LRNWDIGPQQTQVGIQYGVQTVVHEFVLTNTSYTTEVMDAALRIQRGGGTQMTALGIDT 246	
QY	252	ACTEGFSQSGRRPEARLLVWTDGESHGDBELPAALKACBAGRVTTRYGIAVLGHYLR 311	
DB	247	AREEAFTEAGRGVGVQVWVITVDGESHONYRLQEVIDKCEDENIQRFALILGYSRG 306	
QY	312	QROPSSFLREITIASDPDRFPFNTVDEAALTDIVDALGDRIFGLGSHAENESSFGL 371	
DB	307	NLSTEKEVEIKSIASPTKHPFNVSDELALVTIVEALGERIPALEATTDDQQAASFE 366	
QY	372	MSQIGFSTRKKGILFGMVGAYDWGGSVLWLEGGHRLFPFRMALEDFPPALQNHAAYL 431	
DB	367	MSQAGFSAHYSDQWVWMLGAVGVDWGTVMVWXSISIPSDNTRFRHSEKIEPLAAYL 426	
QY	432	GYSVSSMLLGGRLFLSGAPRPHRGKVIAFOLKDGAVRVAQSLQGBQIGSYFSGELC 491	
DB	427	GTYVNSALTPG-VLYIAGQPRYNHTQCVIYKRM-EGREVKVLQRLKGEIQIGSYFGVIT 484	
QY	492	PLDTRDGTDLVLAAPMFLGPONKETGRVYVVLVG-----QQSLTLTQGT----- 538	
DB	485	TIDINRDSFTDLLLVGAPTYMGTEKEGQVYVVALNKTFFQVMSLEPIKQTCSSPLKH 544	
QY	539	-----LOPEPPQDARFGAMGALPDNLNODGADVAGAPLEDHGQALYLYHGTOGVRPH 594	
DB	545	DTCKVLKNEPCGARFGTAIAAVKDLNDGYNDIVIGSPLDDHRAVYIVHGHGNTISK 604	
QY	595	PAQRIAAASHALSYGRSDVGRDLDDGDLVDVAVGAQAAILLSRIVHLTPSLEV 654	
DB	605	YTORIASGGDGEKVFEGSQSVHGBMDLDDGLIDVTIGLGGALFWSRDAEVNVSQMF 664	
QY	655	TPQAISSVQDRCRRRGOEAVCLTAALCFQVTSRTPGRWD-HOFYMRFTASIDWETAGARA 713	
DB	665	MPKSINTQQNCQNKKTICINATICFK--TRLKSIEDIFESSLQVWITLDAQRQISRS 722	
QY	714	AFDGGQRLSPRRLRLSVGNVTCBQLPHVLDTSDYLRPVALTVTTFALDNTTKP--GPVL 771	
DB	723	LFTETHERKMQK--NITIKGSECIKHFYMLDKDPQDSVKVLEF---NFSDESQPLV 777	
QY	772	NEGSPTGIQLVPRSKCGPDNECVTLVLQVNMIDIRGSKAPVVRGGRKVLVSTLE 831	
DB	778	DSNLPNISISYIPTKDCGAKNKCISDLALNVKASIAGDSSPPFVKSRRNDRFTIQLSIK 837	
QY	832	NRKENAYNTLSIIFSRNLHLASLTP-QRESPIKVECAAPSAHARLCSVGHVPVOTGAKV 890	
DB	838	NKDSANTRVALVQYSNITIFAGIEDIQDS-----CRSNHMITCKVGPFLKPEEI 890	
QY	891	TFLEPFSCSSLSQVFGKLTASSDSLERNGLTQENTAQTSAYIQYEPHLLFSSESTLH 950	
DB	891	SFKISFQFNASYLLENATVDVYATSDSEEPETLRDNRGQVTIPVKYEVGLIFVSVFKEH 950	
QY	951	RYEVHPGTPLVPG-----PGPEFKTLRLVQNLGCYVVGSLIISALLPAVAHGN---YF 1001	
DB	951	HVIIAANETIPTAINTTEQIGDEVTLYHRIEKGHFFMPNLTQLIQLYPDVTTAKNTLLYI 1010	

QY	1002	LSLSQVITNNASC-----IVONLTPEPPPPVHBELOHTNRLNGSNTQ 1044	
DB	1011	TTLSH--SQNAICKSSYPVDHLKIGSKSVLPKIKEPTKDTI-----MECDTFS 1058	
QY	1045	QOVVRCHLQOLAKGTBVSGLLRLVHNEFFRAKFKSLTVSTFELGTBEGSVLQUTEAS 1104	
DB	1059	CASINICALAP-SDISQVNVSL--RVKPTTIKASIHSLTLVVKALLRSNSLLIRNDHQ 1115	
QY	1105	RWSESLELVVQTPPILISLWILIGSVLGLLALLVFCMLKGLFPFAHKKIPPEEKREE 1163	
DB	1116	KUETMIKISKEPPPGSVPLWVPLSFAGLLILALLIFALWKAGFF---KRPLKKKMK 1171	
RESULT 8			
IT2A_HUMAN			
ID	IT2A_HUMAN	STANDARD; PRT; 1181 AA.	
AC	P17301;		
DT	01-AUG-1990 (Rel. 15, Created)		
DT	01-AUG-1990 (Rel. 15, Last sequence update)		
DT	25-OCT-2004 (Rel. 45, Last annotation update)		
DE	Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)		
DE	(Collagen receptor) (VLA-2 alpha chain) (CD49b).		
GN	Name=ITGA2;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 30-44.		
RC	TISSUE=Endothelial cells;		
RC	MEDLINE=89308879; PubMed=2545729; DOI=10.1083/jcb.109.1.397;		
RA	Takada Y., Hemler M.E.;		
RT	"The primary structure of the VLA-2/collagen receptor alpha 2 subunit (platelet GP1a): homology to other integrins and the presence of a possible collagen-binding domain.";		
RT	J. Cell Biol. 109:397-407(1989).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Rieder M.J., Amel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,		
RA	Rajkumar N., Toch E.J., Yi Q., Nickerson D.A.;		
RT	"SeattlesNP. NHLBI HL66682 program for genomic applications, UW-		
RT	PHCRC, Seattle, WA (URL: http://pga.gs.washington.edu).";		
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.		
RX	MEDLINE=98019223; PubMed=9353312; DOI=10.1074/jbc.272.45.28512;		
RA	Emley J., King S.B., Bergelson J.M., Liddington R.C.;		
RT	"Crystal structure of the I domain from integrin alpha2beta1.";		
RL	J. Biol. Chem. 272:28512-28517(1997).		
RN	[4]		
RP	VARIANT GLU-534.		
RX	MEDLINE=94043762; PubMed=7901236;		
RA	Santoso S., Kalb R., Walka M., Kiefel V., Mueller-Eckhardt C.,		
RA	Newman P.J.;		
RT	"The human platelet alloantigens Br(a) and Br(b) are associated with a single amino acid polymorphism on glycoprotein Ia (integrin subunit alpha 2).";		
RT	J. Clin. Invest. 92:2427-2432(1993).		
RN	[5]		
RP	VARIANT GLU-534.		
RX	MEDLINE=20206009; PubMed=10744142;		
RA	Kroll H., Gardemann A., Fechter A., Haberbosch W., Santoso S.;		
RT	"The impact of the glycoprotein Ia collagen receptor subunit Al648G gene polymorphism on coronary artery disease and acute myocardial infarction.";		
RL	Thromb. Haemost. 83:392-396(2000).		
CC	-I- FUNCTION: Integrin alpha-2/beta-1 is a receptor for laminin, collagen, collagen C-propeptides, fibronectin and B-cadherin. It recognizes the proline-hydroxylated sequence G-F-P-G-E-R in collagen. It is responsible for adhesion of platelets and other cells to collagens, modulation of collagen and collagenase gene expression, force generation and organization of newly synthesized		


```
Db 310 AVILGLNRLNADTKNLIKAIKAIPTERYFFNVSDAAALKEAGTLGQIFSIET-V 368
Qy 363 ENSSFCLEMSQTCFST--HRLKDGILFGMVGAYDMGSGVLM-LEGHRLFPFRMALEDE 419
Db 369 QGGDNFQWMSQVGSFADYSQNDILMLGAVGAFGWSGTTIVQKTSHGHLIFF-----KQA 423
Qy 420 FPPALQ--NHAAYLVGSVSMILRGRRRLFLSGAPRPHRGKVIKFKKDGAVRVAQSL 477
Db 424 FQILQDRNHSSYLVGSVAA-ISTGESTHFVAGAPRANYTGQIVLVSVNENGMITVIAH 482
Qy 478 QGSGIGYFSGELCPLDTRDGTDLVLLVAAPMLPGQNKETGRVYVY-----LVGQOQL 532
Db 483 RGDQIGSYFGSLVCSVDVDDTITDVLVGAAPYMSDLKKEGRVYLFITKKGILGQHOF 542
Qy 533 LTLQGTLOPEPPQDARFGFAMGALPDNLQDGPADVAGAPLEDHCHQALYHCTQSGVR 592
Db 543 --LEG---PGIENTRFGSAIALSDINMGDFNDVIVGSPLENQNSGAVIYNGHQITP 597
Qy 593 PHPAQRIAAA--SMPHALSFGRSVDGRDLDDGDLVDVAVGAGAAILLSSRPVHLTP 650
Db 598 TKYSQKILGSDGAFRSHLQFGRSLDGYGLNGDSITDVSIGAFQGVVQVLMWSQSIADVAI 657
Qy 651 SLEVTQAIISVQDRCRRRQGEAVCLTAALCFQVTSRTPGRWDHGFYMRFTASLDEWTAG 710
Db 658 EASFTPEKITLVNKAQ-----IILKLCFSAPK-PTKQNNQVAIVYNTILDA---- 704
Qy 711 ARAAFDGSQGLSPRL-----RLSVGNV-----TCEQLHFWLDTSDYLPVALTVT 758
Db 705 -----DGFSSRVTSRGLFKENNERCLQKNVVAQAQSCPEHIIYIQBPSVMSLDLRVD 759
Qy 759 FALDNTTKPG--PVLNKGSPTSIQKLVFSSKDCPDNECVTDILVQVMDIRGSRKAPV 816
Db 760 ISLEN-----PQTSAPALEAYSETAKVFSIPFKDCGEDGLCISDLVDYR-QIPAAQEQPFI 815
Qy 817 VRGRRKVLVSTTLNENKENAYNTLSIIIFSRNLHLASLTPQRESPI---KVBC-AAPSA 872
Db 816 VSNQNKRLTFSVTLNKNKREAYNTGIVVDFSENLFASFS-----LPVDGTEVTCQVAASQ 871
Qy 873 HARLCSVGHVPVQFGAKVTLLEPFSCSILSQVFGKLTASSDSLSRNGTLQENTQAITS 932
Db 872 KSVACDVGYPALKRQEQVTTINDFNLQNLQNASLSFQALSQESQENKA--DNLVNLK 929
Qy 933 AYIQYEPHLLFSBSTLHRYEVHPYGTLP-----VGPGPEFTTLRVQNLGCVVWSGLI 986
Db 930 IPLYLDAIEHLTRSTNINFEISSDGNVPSIVHSFEDVGPKFIFSLKV-TTGSVPVSMAT 988
Qy 987 ISALLPAVAGGNYFLSLSQVITNNASCIQNLTEPP-----GPPVHPEELQHTNRL 1038
Db 989 VIIHIPQYTKBNPLMYLTGVQTDKAGDISCNADINPLKIGQTSVSSVSFKSENPRHTKEL 1048
Qy 1039 NGSNTQCVVCHLQOLAKTEVSGVLLRLVHNEFFPRAKFSLTVVSTPELCTEESVL 1098
Db 1049 NCRFASCSNTVCLWKDVMHKGEPVNVTTIRWNGTFASSTFTQVLTAAABINFTNPEIY 1108
Qy 1099 QLTEASRWSLLEVVQTRPLISLMI-----LIGSVLGLLILLALLVFLCW 1145
Db 1109 -----VIEDNTVTPLMTMKPDKAEVPTGVIIIGSIIAGILLALLVALVILW 1154
Qy 1146 KLGFPAHK-----KIPBE 1158
Db 1155 KLGFPRKRYEKWTKNPDE 1172
```

RESULT 9

```
ITA2 BOVIN STANDARD; PRT; 1170 AA.
AC P53710;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).
```

```
GN Name=ITGA2;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94193647; PubMed=7511592;
RA Kanata T., Puzon W., Takada Y.;
RT Identification of putative ligand binding sites within I domain of
RT integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";
RL J. Biol. Chem. 269:9659-9663(1994).
CC -I- FUNCTION: Integrin alpha-2/beta-1 is a receptor for laminin,
CC collagen, collagen C-propeptides, fibronectin and E-cadherin. It
CC recognizes the proline-hydroxylated sequence G-P-P-G-E-R in
CC collagen. It is responsible for adhesion of platelets and other
CC cells to collagens, modulation of collagen and collagenase gene
CC expression, force generation and organization of newly synthesized
CC extracellular matrix.
CC -I- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
CC associates with beta-1.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -I- SIMILARITY: Belongs to the integrin alpha chain family.
CC -I- SIMILARITY: Contains 7 FG-GAP repeats.
CC -I- SIMILARITY: Contains 1 VWFA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L25886; AAB59255.1; -.
CC PIR; I45914; I45914.
CC HSSP; P17301; IAOX.
CC InterPro; IPR000413; Integrin_alpha.
CC InterPro; IPR002035; VWF_A.
CC Pfam; PF01839; FG-GAP; 2.
CC Pfam; PF00357; Integrin_alpha; 1.
CC Pfam; PF00092; VWA; 1.
CC PRINTS; PR01185; INTEGRINA.
CC PRINTS; PR00453; VWFADOMAIN.
CC SMART; SM00191; Int_alpha; 5.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
CC PROSITE; PS50234; VWFA; 1.
CC Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;
CC Polymorphism; Receptor; Repeat; Signal; Transmembrane.
CC NON TER 1 1
CC SIGNAL <1 18 Potential.
CC CHAIN 19 1170 Integrin alpha-2.
CC DOMAIN 19 1121 Extracellular (Potential).
CC TRANSMEM 1122 1143 Potential.
CC DOMAIN 1144 1170 Cytoplasmic (Potential).
CC REPEAT 34 92 FG-GAP 1.
CC REPEAT 93 92 FG-GAP 2.
CC REPEAT 177 367 VWFA.
CC REPEAT 368 422 FG-GAP 3.
CC REPEAT 423 475 FG-GAP 4.
CC REPEAT 477 538 FG-GAP 5.
CC REPEAT 540 599 FG-GAP 6.
CC REPEAT 604 656 FG-GAP 7.
CC CA_BIND 488 496 Potential.
CC CA_BIND 552 560 Potential.
CC CA_BIND 616 624 Potential.
CC SITE 472 474 Cell attachment site (Potential).
CC SITE 1146 1150 GFPKR motif.
CC DISULFID 72 81 By similarity.
```

```

FT DISULFID 669 726 By similarity.
FT DISULFID 778 784 By similarity.
FT DISULFID 854 865 By similarity.
FT DISULFID 1008 1039 By similarity.
FT DISULFID 1044 1049 By similarity.
FT CARBOHYD 94 94 N-linked (GlcNAc...) (potential).
FT CARBOHYD 101 101 N-linked (GlcNAc...) (potential).
FT CARBOHYD 332 332 N-linked (GlcNAc...) (potential).
FT CARBOHYD 421 421 N-linked (GlcNAc...) (potential).
FT CARBOHYD 449 449 N-linked (GlcNAc...) (potential).
FT CARBOHYD 464 464 N-linked (GlcNAc...) (potential).
FT CARBOHYD 688 688 N-linked (GlcNAc...) (potential).
FT CARBOHYD 748 748 N-linked (GlcNAc...) (potential).
FT CARBOHYD 945 945 N-linked (GlcNAc...) (potential).
FT CARBOHYD 1063 1063 N-linked (GlcNAc...) (potential).
FT CARBOHYD 1070 1070 N-linked (GlcNAc...) (potential).
FT VARIANT 580 580 G -> V.
FT VARIANT 588 588 R -> K.
FT VARIANT 725 725 R -> S.
SQ SEQUENCE 1170 AA; 128929 MW; ECEPICSF2448FB1 CRC64;

Query Match
Best Local Similarity 33.9%; Pred. No. 6.1e-116;
Matches 411; Conservative 235; Mismatches 465; Indels 102; Gaps 30;

9 LFLPLVLTGL---CSPFNLDLHPRLPFGPEAEFGYSLVQHVGCGQRMLVCAWDGP 65
2 LQLVLVSQGLNCCVAYNGLPKAKIFSGSPSEQFGYAVQOFTNPKGNMLLVGSPWSGF 61
66 SGDRRGDVRCPVCGAHPACAKGHL--GDYQLGNSSHAPVNMHGMISLLETGDDGGFMAC 124
62 PKRMGDVYKCPV--DLSTTTCYKLNLTSTMSNVTENKMSLGLTLTRNVGTGGFLTC 120
125 APLWSRACSSVFSFGICARVDASFPQGS LAPTAQRCPTVMVIVLDGNSNIYPMSEV 184
121 GPLWAQCGSQYTTGVCSDVSPDFQLRTSPAPAVQTCPSFIDVVVCDENSIYPWDVA 180
185 QTFRLRLVGKLFIDPEQIQVLGVYGESPVHWSLGDPRFKYKEEVVRAAKNLSRREGRETK 244
181 KNFLEKFEVQGLDGTFTKTMGLIQYANNPRVFNLTFSKDNEMIKATSTQFYGGGLTN 240
245 TQAQIMVACTEGFSQSHGGRPEARLLVWVTDGSHDGEELPAALKAACEAGRVTRYGIAY 304
241 TFAQIAYARDATYSTAAGRGATKVMVWVTDGSHDGSKLKAVIDQCNKNIILRFGIAY 300
305 LGHYLRQRDPSSFLRIRTIASDPDRFFNFVDEAALTDIVDALGDRIFGLRGSNAEN 364
301 LGYLNRNALDTKNLKEIKAIASIPTEHFNVSDEADLLEKACTIGEIQIFSIEST-VQG 359
365 ESSFGLMSQIGFSTHRLKDG--ILFGVGYAYDGGVSLM--LEGHRLFPFRMALEDEFP 421
360 GDNFQEMSQVGFSAEYSPQNNILMLGAVGAYDMSGTVVQKTPHGLIFS-----KQAFE 414
422 PALQ--NHAAYLVGYSVSMLLGRRRLFLSGAPRFRHGVKVIAPQLKDKGAVRVAOSLOG 479
415 QILQDRNHSSYLVGYSVAS--ISTGNSVHFVAGAPRANTYTGQVLVSVNENGVTVIQSQRG 473
480 EQIGSYFSGELCPDTRDGTDTLLVAAPFLGPQNKETGRVTVYLVGQOSLLTLQGTLL 539
474 DQIGSYFSGVLCAVDVKNKDTITDVLVAGAPMYMNDLKEEGRVYLFITKILNWHQFLE 533
540 QPEPPQDARFGAMGALPDNLQDGFADVAGAPLEDGHQGLYLYHGTQSGVRPHPAQRI 599
534 GPNGLENAREGSAATAALSDINMDGFNDVIVGSPLENQNSGAVIYNGHEGMIRLRYSQKI 593
600 AAASMPHA--LSYFGRSGVGLDGLDGDLDVAVACAGAAILLSSRPVHLTPLEVTTPQ 657
594 LGSDFRASFSLHYFGRSLDGYDGLGDSITDVSVGAFQGVVQLVQSADIADVSADSTPK 653
658 AISVQVQDCRRRQGEAVCLTAALCFQVTSRTPGGRWDHGFYVNRFTASLDEWTAG----ARA 713
654 KITLLNKAEE-----IKLKLCSAKFR--PTNNQNVAVIYNTIDEDQFSSRSVIRG 704

```

```

Qy 714 AFDGSGORLSRRLRSVGNVTCQQLHFLVLDTSYLRLPVALTVTTFALDNTTKPG--PVL 771
Db 705 LFKENNERCLOKTMIVSQAQ-RCSEYIIHQEPIISPLNLCNLSLEN---EGTNPAL 760
Qy 772 NEGSPSTIQKLVPFKDCGPDNECVTDLVQVNDMDIRGSRKAPFVWGRGRKRVLSVTLE 831
Db 761 EAYSETVVFISIPFKDCGDDGVCISDLVLNV-QQLPATQQOPIFVSNQNKRLTFSVLK 819
Qy 832 NRKENAYNTSLSIIFSRLNLHLASLTPORESPI---KVEC-AAPSAHARLCSVGHVPQTG 887
Db 820 NKKEAYNTSIVVDVDFSENLPFASWS---MPVDGTEVTCQIASSQKSVTCNVGYPAKSK 875
Qy 888 AKYTELEFEFSCSSLSLQSVFGKLTASSDSLSRNGTLQENTAQTSAYIQIYEPHLLFSSES 947
Db 876 QQVTFINFDNQLNQNASISPRALSEOEEN--MADNSVNLKULLYDAEIHITKST 933
Qy 948 TLHRYE-----VHPYGTLPVGPBPFKTLRLVQNLGCVYVSGLIISALLPAVAHG 997
Db 934 NINFYEVSLDGNVSVVHSFEDI---GPKFIFSIKV-TTGSVPVSMASVIIHIPOYTKD 988
Qy 998 GNVFLSLSQVITNNA---SCIVQNLTPEPPGPVHP-----BELQHTNRLNG 1040
Db 989 KNPLMYLTGVHTDQAGDISCEAE-----INPLKIGOTSSSVFKNSEFRHKEKELNC 1039
Qy 1041 SNTQCQVVRCHLGQAKGTEVSVGLLRLVHNEPRRAKEKSLTVVSTFELGTEEGSVLQL 1100
Db 1040 RTASCSNIMCLRDLOVKGEYFLNVSTRIWNGTFAASTFQVLTAAEIDTYNPQIYVI 1099
Qy 1101 TEASRWSESLLEVQVTRP---ILISWILGSLVGLLLALLVFLCWLKLGFFPAHK---- 1153
Db 1100 EE-----NTVPIPLTMKPEKVEVPTGVIVGSIAGILLALLVAILMLKLGFFRKYKEM 1155
Qy 1154 -KJPEEKREKL 1165
Db 1156 AKNPDETDETEL 1168

RESULT 10
ITAZ MOUSE
ID ITAZ2 MOUSE STANDARD; PRT; 1178 AA.
AC Q62469; Q62163;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
GN Name=Itga2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
PI [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=94363406; PubMed=8081889;
RA Edelman J.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,
RA Damjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;
RT "The mouse VLA-2 homologue supports collagen and laminin adhesion but
RT not virus binding.";
RL Cell Adhes. Commun. 2:131-143(1994).
RN [2]
RP SEQUENCE OF 450-1178 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=94355691; PubMed=7521231;
RA Wu J.B., Santoro S.A.;
RT "Complex patterns of expression suggest extensive roles for the alpha
RL Dev. Dyn. 199:292-314(1994).
CC -I- FUNCTION: Integrin alpha-2/beta-1 is a collagen receptor, being
CC responsible for adhesion of platelets and other cells to
CC collagens, modulation of collagen and collagenase gene expression,
CC force generation and organization of newly synthesized
CC extracellular matrix. It is also a receptor for laminins, collagen

```

CC C-propeptides and E-cadherin. Mice homozygous for a null mutation
 CC in the alpha-2 die very early in embryogenesis.
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
 CC associates with beta-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
 CC with I-domains do not undergo protease cleavage.
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -1- SIMILARITY: Contains 1 VWFA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z29987; CAA82877.1; -;
 CC EMBL; X75427; CAA53178.1; -;
 CC PIR; S44142; S44142.
 CC HSP; P17301; IAOX.
 CC MG; MGI:96600; Itga2.
 CC InterPro; IPR000413; Integrin_alpha.
 CC InterPro; IPR002035; VWF_A.
 CC Pfam; PF01839; FG-GAP; 2.
 CC Pfam; PF00357; Integrin_alpha; 1.
 CC Pfam; PF00032; VWA; 1.
 CC PRINTS; PRO1185; INTEGRINA.
 CC PRINTS; PRO0453; VWFADOMAIN.
 CC SMART; SM00191; Int alpha; 5.
 CC SMART; SM00327; VWA; 1.
 CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 CC PROSITE; PS0234; VWA; 1.
 CC Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;
 KW Receptor; Signal; Transmembrane.
 FT SIGNAL 1 26
 FT CHAIN 27 1178
 FT DOMAIN 27 1129 Integrin alpha-2.
 FT TRANSEM 1130 1151 Extracellular (Potential).
 FT DOMAIN 1152 1178 Potential.
 FT REPEAT 42 100 Cyttoplasmic (Potential).
 FT REPEAT 101 ? FG-GAP 1.
 FT REPEAT 185 ? VWA.
 FT DOMAIN 185 375 VWA.
 FT REPEAT 376 430 FG-GAP 3.
 FT REPEAT 431 483 FG-GAP 4.
 FT REPEAT 485 546 FG-GAP 5.
 FT REPEAT 548 607 FG-GAP 6.
 FT REPEAT 612 664 FG-GAP 7.
 FT CA_BIND 496 504 Potential.
 FT CA_BIND 560 568 Potential.
 FT CA_BIND 624 632 Potential.
 FT SITE 480 482 Cell attachment site (Potential).
 FT SITE 1154 1158 GFPR motif.
 FT DISULFID 80 89 By similarity.
 FT DISULFID 677 734 By similarity.
 FT DISULFID 786 792 By similarity.
 FT DISULFID 862 873 By similarity.
 FT DISULFID 1016 1047 By similarity.
 FT DISULFID 1052 1057 By similarity.
 FT CARBOHYD 102 109 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 109 109 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 429 429 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 457 457 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 472 472 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 696 696 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1054 1054 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1071 1071 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1078 1078 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 1178 AA; 128926 MW; 1P194B9C0240F465 CRC64;

Query Match

28.6%; Score 1744; DB 1; Length 1178;

Best Local Similarity 35.3%; Pred. No. 3.2e-115;
 Matches 420; Conservative 221; Mismatches 479; Indels 70; Gaps 31;
 QY 9 LFLPLVFLT-GL--CSPFNLDDEHHPRLFPQPEAERGYSVLQHVGGQRMVNLVGPMDGP 65
 DB 10 LLLQLMLVQGIILCLAYNVGIPGAKIFGPGSEQFGYSVQQLTNPQGNWLLVGSWMSGF 69
 QY 66 SGDRRGDVYRCPVGGAHNAPCAKGHIL-GDYQLGNSSHPAVNMHLGMSLLETDGCGFMAC 124
 DB 70 PENMGDVYKCPV-DLPTATCEKLNQNSASISNVTETKTNMSLGLTLTENPGTGGFLTC 128
 QY 125 APWMSRACSSVSSGICARVDASPOQSLAPTACRCPTVMYDVIIVLQGSNSIYPWSEV 184
 DB 129 GPLWAHQCGNYATGICSDVSPDFQFLTSFSPAQCPSLVDVVVVCDESNIYPWEAV 188
 QY 185 QPLRLVGLKFLDPQIIOVLVOYCESPVHWSLGDFTKEBVVPAAKNLSRREGTEK 244
 DB 189 KNFLVPFVTGLDIGPKKTQVALIQYANEPRIIPLNLDFTKEDMVOATSETROHGDLTN 248
 QY 245 TAAIMVACTEGFSQSHGGRPEAARLLVVVTGESHGDELPALAKACAGRVTRYGIIV 304
 DB 249 TPAIEFARDYAYQSOTSGRPGATKVMVVVTGESHGSKLKTIVIOQNDDELIRGLIV 308
 QY 305 LGHYLRQRDPSSFLREIRTIASDPDERFFNVYDDEAALTVIDALGDRIFLEGSHAEN 364
 DB 309 LGYLNRNALTQNLKEIKAIASPTPTERYFFNVADAEALKEAGTLGEQIFSTEGT-VQG 367
 QY 365 ESSFGLEMSQIGFSTHRL--KQILGMYGVDGWSVLWLEGHR--LPPPRMALEDEF 420
 DB 368 GDNFQEMAQVGSADYAFONDILMLGAVGAFDWSGTLV-QETSHKPVFPF-----KQAF 421
 QY 421 PPAIQ--NHAAYLGYSVSMLLRGRRFLSGAPRHRGKVIATFOLKXKDGAVRVASQLQ 478
 DB 422 DQVLQDRNHSSFLGYSVAISTEDGVH-FVAGAPRANYTGOIVLYSVNKGNTVTIQSHR 480
 QY 479 GEIGSYFSGELCPLDTRDGTVDLLVAAPMPLGPONKETGRVYVYLVGQSSLLTQGT 538
 DB 481 GDQIGSYFGSVLCSDVDVDKDTITDVLVGAPTYMNDLKEBEGKVYLTITKGLNQHQFL 540
 QY 539 LOPEPQDARRGFAMGALPDINODGFADVAVGAPLEDGHQGLYLVHGTSGVRPHPAOR 598
 DB 541 EGPEGTGNARFGSAIAALSDINMDGPDVIVGSPVENENSGAVIYNGHQGTIRTKYSQK 600
 QY 599 IAAA--SMPHALSYFGRSVDGRDLDDGLVDVAVGAQGAAILLSRPIVHLTPSLEVTP 656
 DB 601 ILGNGAPRRHLQFFGRSLDGYGLNGDSITDVISIGALQVQLWSQSADVAIEALFTP 660
 QY 657 QAISVVQDRCRRRQGBAVCLTAALCFQVTSRTPRWDHOFYMRFTASLDEWTAGARAAD 716
 DB 661 DKITLLNKDAK-----ITLKLCPFRABFRPAGQ-NNQVAILFNMTLDADGHSRVTSR 711
 QY 717 GSGQRLSPRLR--LSVGNV-TCEQLHFHVLDTSYLRPVALTVTFTALDNTTKPG--PVL 771
 DB 712 GVFPRESERPLQKNVNVNEVQKSEHHISIQKPSDVNPLDRLVDISLEN---PGTSPAL 769
 QY 772 NEGSPTSIOKLVPFSKDCGPDNECVTDVLQVNMIDIGSRKAPFVVRGRRKVLVSTTLE 831
 DB 769 EAYSEIVKFSIPFYKECGSDGICISDLILDV-QQLPAIQTSQSFIVSNQNKRLTFSVLK 827
 QY 832 NRKENAYNTSLSIIPSRNLHLASLTPORESPI---KVECAAPSAHARL-CSVGHVPQTG 887
 DB 828 NRGESAYNTVLAEPSENLFASFSS---MPVDGTEVTECVGSSQKSVTCVGVYPALKSE 883
 QY 888 AKYTELEPEFSKSSLLSGLVFGKLTASSDSLENGTLQENTAQTSAYIOVEPHLLSSES 947
 DB 884 QQVTFITNFDNFNLQNLQNAIINFQAFSESQETNKA--DNSVSLTIPLLYDAELHLTRST 941
 QY 948 TLHRYEVHPYGTLP-----VGPGEFRTKTLRQNLGCVYVSGLIISALLPAVAGHNGYF 1001
 DB 942 NINFYISSDENAPSVIKSVEDIGPKFISLKV-TAGSAPVSNALVTIHIPOYTKENPL 1000
 QY 1002 LSLSQVITNNA---SCIT--VQNLTEP---PGPPVHPPELOHTNRLNGSNTOCQVVRCHLG 1053

Db 1001 LYLGTGTDQAGDISCTAEINPLKLPHTAPSVSFKNENPRHTKELDCRTTSCSNITCWLK 1060
 Qy 1054 QLAGGTVEVGLRLVHNEFFRAKFKSLTVVTFELGTBEGSVQLQTEASRWSESLLEV 1113
 Db 1061 DLHMAEYFNVITRVNRTFAASTFTQVLTAAAEIDTHNPO-LFVIEENAVTIPLMIM 1119
 Qy 1114 VQTRPILISLWILIGSVLGLLLALLVFLCWLKLGFF--AHKKI---PEE 1158
 Db 1120 KPTEKAEPVGTGVIIGSIAGILLLLAMTAGLWKLGLGFFKRYKKMGQNPDE 1169

RESULT 11
 Q6P1C7
 ID Q6P1C7 PRELIMINARY; PRT; 1178 AA.
 AC Q6P1C7
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Integrin alpha 2.
 GN Name:Itga2;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Eye;
 RX MEDLINE=2388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Eye;
 RA Strausberg R.;
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
 DR EMBL; BC065139; AAH65139.1; -.
 DR HSSP; P17301; 1A0X.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR InterPro; IPR00413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 2.
 DR Pfam; PF00357; Integrin_alpha; 1.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS0234; VWFA; 1.
 KW Cell adhesion; Integrin; Transmembrane.
 SQ SEQUENCE 1178 AA; 128954 MW; 62FAEA820242A9B6 CRC64;

Query Match 28.6%; Score 1743.5; DB 2; Length 1178;
 Best Local Similarity 35.2%; Pred. No. 3.5e-115;
 Matches 417; Conservative 222; Mismatches 477; Indels 69; Gaps 30;
 Qy 13 LVFLTGL--CSPNLDHHRPLRFPGPPEAEFGYSVLQHVGGGORMWLVGAPWDGPGSDRR 70
 Db 15 LMLVQGLNCLAYNVLPGAKIFSGPSSEQFGYSVQQLTNPQGNWLLVGSFWSGFENRM 74
 Qy 71 GDVYRCPVGAHNAAPCAKHL--GDYQLGNSSHAPVNMHLGMSLLETDDGGDFMACAPLWS 129
 Db 75 GDVYKCPV-DLPTATCEKLNQNSASISNVTEIKTNMSLGLTLTRNPGTGFLTCGFLWA 133
 Qy 130 RACGSSVFSSGICARVDASPOGSLAPTAQRCPYMDVIVLDGNSIYPMWSVQVFLR 189
 Db 134 HQCGNQYATGICSDSPDFQTLTSPFPAVQACPSLDVVVVVCDSEMSIYPMWAKNPLV 193
 Qy 190 RLVGKLFIDPEQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKAI 249
 Db 194 KFTVGLDIGPKTQVALIQVANEPRIFNLNDETREDVMQATSETRQHGDLTNTFRAI 253
 Qy 250 MVACTEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHL 309
 Db 254 EFARDVAYSQTSGRPCATKVMVWVTDGSHDGSKLTVIQCCNDDDEILRFGIAVLGYN 313
 Qy 310 RRQDPSSFLREIRTASDPDERFFENVYDDEALTDIVDALGDRIFGEGSHAENESSFG 369
 Db 314 RNALDTKNLKEIKATASTPTEYFFNVDAEALKEKAGTLGEOIFSIETG-VGGGDNFQ 372
 Qy 370 LEMSQIGFSFTRL--KXGILFGVMGAYDMGSGVLWLEGGHR--LFPFPRMALEDFFPALQ 425
 Db 373 MENAQVGSADYAPONDILMLGAVGAFDWSGTLV-QETSHKPVIFP-----KQAFDQVLQ 426
 Qy 426 --NHAAYLVGSVSMILRGRRRLFLSGAPRFRHGRKVIAFOLKKGAVRVQAISLQSGQIG 483
 Db 427 DRNHSSFLGYSVAAISTEDGVH-FVAGAPRANTGQVLVSVNKGQNVTVQSHRGDQIG 485
 Qy 484 SYGSELCLPDTDRDGTITDVLVAAPFLGPONKETGRVVVYLVQQSLLTLOQTLOPEP 543
 Db 486 SYFGSVLCSVDVDKDTITDVLVGAPYMDLKKBEKGVLYLTITKILNQHOHFLGEPG 545
 Qy 544 PQDARFGMAGALPDNLNODGFADVAVGAPLEDHGQALYLYHGTSQSVRPHAPQRIAAA- 602
 Db 546 TGNARFGSAALSDINMGDFNDVIGSPVENENSGAVIYNGHQITIRKYSQKILGSN 605
 Qy 603 -SMPHALSFGRSVDGRDLDDGLVDVAVGAAGAAIILSSRPVHLVTPSLEVTPQAISV 661
 Db 606 GAFRRHLQFGRSLDGYDLNGDSITDVSIGALQVQLWSQSIADVAIEALFDPDKITL 665
 Qy 662 VQRDCRRRGQEAIVCLFAALCFQVTSRTPGRWDHQFYMRFTASLDDEWTAGARAAFDGSGOR 721
 Db 666 LNKDAK-----ITLKLCFRAEFEPAGQ-NNQVAILFNMTLDADGHSSRVSRTSGVPRE 716
 Qy 722 LSPRRLR--LSVGNV-TCEQLHFLVLDTSVLRPVALTVTFALDNTTKPG--PVLNKGSP 776
 Db 717 NSERFLQNMVMEVQKCEHHISIQKPSDVVNPDLRVDISLEN---PGTSPALEAYSE 773
 Qy 777 TSIQKLVFPKDCPDNECVTDLVQVNMIDIRGSRKAPFVVRGGRKRVLTSTLTLENKEN 836
 Db 774 TVKVFSIPFYKEGSDGICISDILDV-QQLPAIQTSFIVSNQNKLTFTSVILKNRGES 832
 Qy 837 AYNTSLSIIFSRNLHLASLPQRESPI---KVECAAPSAHARL-CSVCHPVFOGAKVTF 892
 Db 833 AYNTVLAEEFSENLFASFSS----MPVDGTEVTCEVSSQKSVTCGVGPAKSEQQVTF 888
 Qy 893 LLEFEFSCSLLSQVFGKLTASDSLERNGTLOQNTAQTSAIIOYEPHLLFSSSTLHRY 952
 Db 889 TINPDLNLQNLQAAINFOAFSESQNTKA--DNSVSLTIPLYDLIELHLTRSTNIFY 946
 Qy 953 EVHPYGTLP-----VGPGEFKTTLRVQNLGCVVYVSGLIISALLPAVAGHGNFYLSLQ 1006
 Db 947 EISSDENAPSVIKSVEDIGPKFTFSLKV-TAGSAPVSMALVTIHPQVTEKKNPLLYLTG 1005
 Qy 1007 VITNNA---SCI---VQNLTEP---PGPPVHPEELQHTNRLNGSNTQCVVRCHLGQLAKG 1058

```
Db 1006 IQTDQAGDISCTABINFLKPLHTAPSVFKNENPRHTKELDCRTSCSNITCWLKDLHMK 1065
Qy 1059 TEVSUGLLRLVHNEFFRRKFKSLTVVSTPELGTGEGSVLQLTEASRWSLSLLEVVQTRP 1118
Db 1066 AEFYINVTTRVWNTFAASTFTQVQLTAAABIDTHNQ-LFVIEENAVTIPLMIMKTEK 1124
Qy 1119 ILISLWILGSLVGLGLLLALLVCLMKLGF--AHKKI---PBE 1158
Db 1125 AEPVGTGVIIGSIIAGILLALLAMTAGLKLGFKQYKKMGQNPDE 1169

RESULT 12
ID Q8WY18 PRELIMINARY; PRT; 823 AA.
AC Q8WY18;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MSTP018.
OS Homo sapiens (Human).
GN Homo sapiens.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aorta;
RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
RA Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,
RA Sun Y.H., Jiang B.Q., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
RA Gao R.L., Qiang B.Q., Yuan J.G., Liu C.C., Zhao M.S., Hui R.T.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AF111999; AALJ9001.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PR01185; INTEGRIN.
DR SMART; SM00191; Int alpha; 4.
DR KEGG; K04401; Integrin; Transmembrane.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 823 AA; 92672 MW; D84E78079DCD4925 CRC64;

Query Match 24.4%; Score 1487.5; DB 2; Length 823;
Best Local Similarity 39.6%; Pred. No. 4e-97;
Matches 327; Conservative 151; Mismatches 308; Indels 39; Gaps 13;

Qy 372 MSQIGFSTRLKDGILFGMVGAYDVGSGVLWLEGGHRLFPFMALEDEFFPALONHAYL 431
Db 1 MSQIGFSSHVVEDGVLGAVYDNGAVLAKETSGAGKVIPLRESYLFEPBELKNHGAYL 60
Qy 432 GYSVSSMLLRGGRFLSGAPFRHRGKVIAFQLKDCGAVRAVQASLOEQIGSYFGSELG 491
Db 61 GYTVTSVSSRQGRVYVAGAPFRNHTGKVLFTWHNRSLSLTIHQAMRQQIGSYFGSEIT 120
Qy 492 PLDTDRGDTTVLLVAPMFQPNKETGRVYVYLVGQQSLTLTQGLTQPPPP-QDARFG 550
Db 121 SVDIDGQDVTDLVVGAPMYFN-EGRERKGVYVEL-RQNRFFVYNGTLKDSHSYQNAF 178
Qy 551 FAMGALPDNODGADVAVGAPLEDHCGALYLVHGTQSGVRPAPORIAASMPHALSY 610
Db 179 SSIASVRDLNODSNDVVVGLPNDHAGAIYIFHGFSGSLTKTPKORITASELATGLQY 238
Qy 611 FGRSDGRDLDDGDLVDVAVGACAAILLSSRPVHLTPSLVTPQAISVVQDCRRRG 670
Db 239 FGCSIHGQLDNEGLDLDLAVGALGNVILWSRPVQVNASLHFEPSKINIFHDCRKG 298
Qy 671 QEAVCLTAAALCFQVTSRTPTGRWDHQFYMYRFTASLDIEWTAGARAAPDGGQRLSPRLRLS 730
```

```
Db 299 RDATCAAAFLCFTPIIFLAPHFOQTITVIGIRYNATMDERRYPTRAHLDEGGDRFTNRAVL 358
Qy 731 VGNVTCBQLHFHLDTSYLRLPVALVTTFALDNTTKPGVNLGSGSPSIQKLVFPSKDCG 790
Db 359 SQGLERINFLVLDIADYVKPVTFSVYSLEDDPH-GPMLDDGWNFTTILVSVFPWNGCN 417
Qy 791 PDNECVTDLVQLVNDI-----RGRK-----APFVVRGGRKRLVST 828
Db 418 EDEHCVPLVDARSDLPTAMEYQVRLRKPQADCSAYTLSPDTTTFIESTQRVAVEA 477
Qy 829 TLENKRNAYNTSLSIIFSRNLHLASLTQRESPIKVECAAPS--AHARLCSVGHVPFQT 886
Db 478 TLENKRNAYSTVLNISQSANLQFASLIQKEDSGSIECVNEERRLQKQVCNVSYPFRA 537
Qy 887 GAKVTLLPEFSCSSLSOVFGKLTASSDLSRNGTLOENTATQTSAYIQEYEPHLLFSSE 946
Db 538 KAKVAFRLDFEFSKSLFLHLEIELAAGSDSNERDSTKEDNVAPLPHLYEADVLFTRS 597
Qy 947 STLHRYEVHPYGTLP--VGPGPEFKTTLRVQNLGCVVSGLIISALLPAVAHGNVPLSL 1004
Db 598 SSLSHYEVKLNSSLERYDYGIGPPFCIFRIQNLGLFPFHGMWMMKITIPIATRSNRLKL 657
Qy 1005 SQVITN--NASC-IVQNLTEPPGPPVHPBELQHTNRLNGSNTQCVVRCHLQGLAKTEV 1061
Db 658 RDLFTDEANTSCNIGWNGSTRETPVE-EDLRAPQLNHSNDVSVINCNI-RLVPNQEI 715
Qy 1062 SVGLLRLLVHNEFFRRKFKSLTVVSTFELGTGEGSVLQLTEASRWSLSLLEVVQTRPIL 1121
Db 716 NFHLGLNMLRSALKYKSMKIMVNAALQRPSPFIFREEDPSRQIVFEISKQEDMQV 775
Qy 1122 SLMLIGSVLGGLLALLVCLMKLGFPAHKKIPEEKREEKLE 1166
Db 776 PIWIVGSTUGLLALLVCLMKLGFPPRSAR---RRREPGLD 816

RESULT 13
Q8CE84 PRELIMINARY; PRT; 823 AA.
ID Q8CE84;
AC Q8CE84;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
DE library, clone:4732459H24 product:MSTP018 homolog.
GN Name=Itgall;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RA "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/350555500;
RA RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/350555500;
RA RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RP SEQUENCE FROM N.A.
```


[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RL prepare full-length cDNA libraries for rapid discovery of new genes.";
RN Genome Res. 10:1617-1630(2000).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yanamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Okazaki Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito K., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tegami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AK037222; BAC29761.1; -.
DR MGD; MGI:2153482; Itgal0.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005515; P:protein binding; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF00357; Integrin_alpha; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN_1.
KW Cell adhesion; Integrin; Transmembrane.
FT NON_TER
SQ SEQUENCE 288 AA; 31949 MW; 3f6200f9c9475e85 CRC64;
Query Match 20.8%; Score 1273; DB 2; Length 288;
Best Local Similarity 85.8%; Pred. No. 1.7e-82;
Matches 247; Conservative 20; Mismatches 21; Indels 0; Gaps 0;
Qy 980 GHPVQTGAKVTFLEPFSCSSLLSQVFGKLTASSDSLENGTLQNTAQTASYIQEP 939
Db 1 GHPVQTGAKVTFLEPFSCSSLLSQVFGKLTASSDSLENGTLQNTAQTASYIRYP 60
Qy 940 HLLPSSSTLHRYEHPVGTLPVGPGEFTKTLRVQNLGCVVSGLLIISALLPAVAGGN 999
Db 61 HLVSSTLHRYEHPVGTLPVGPGEFTKTLRVQNLGCVVSGLLIISALLPAVAGGN 120
Qy 1000 YFLSLSQVITNASCIVQNLTPRPPVHPPELOHTNRLNGSNQCQVVRCHLGLAKGT 1059
Db 121 YFLSLSQVITNASCIVQNLTPRPPVHPPELOHTNRLNGSNQCQVVRCHLGLAKGT 180
Qy 1060 EVSVGLLRLVNEPFRRAKFKSLTVSTFELGTREGSVLQLTASRSVSESLLEVQTRPI 1119
Db 181 EISVRLRLVHNEPFRRAKFKSVTVSTFELGTREGSVLQLTASRSVSESLLEVQTRPI 240

Qy 1120 LISLWILIGSVLGGLLALLVFLWKLGFFAHKIPBEKEKLEQ 1167
Db 241 LISLWILIGSVLGGLLALLVFLWKLGFTRKKIPKEESEKLEQ 288
RESULT 15
ITAD RAT
ID ITAD RAT STANDARD; PRT; 1161 AA.
AC OSQVE7;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-D precursor.
GN Name=Itgad;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA O'Brien M.M., Vandervieren M., Kilgannon P.D., Dietach G.,
RA Gallatin W.M.;
RT "Cloning of rat alpha D, a novel beta 2 integrin.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and
CC VCAM1. May play a role in the atherosclerotic process such as
CC clearing lipoproteins from plaques and in phagocytosis of blood-
CC borne pathogens, particulate matter, and senescent erythrocytes
CC from the blood (By similarity).
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D
CC associates with beta-2 (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VWFA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF021334; AAF21241.1; -.
DR HSP; P11215; 1BHQ.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWFA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00234; VWFA; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 1161 Integrin alpha-D.
FT DOMAIN 20 1100 Extracellular (Potential).
FT TRANSMEM 1101 1121 Potential.
FT DOMAIN 1122 1161 Cytoplasmic (Potential).
FT REPEAT 34 87 FG-GAP 1.
FT REPEAT 88 7 FG-GAP 2.
FT DOMAIN 152 334 VWFA.
FT REPEAT 352 402 FG-GAP 3.
FT REPEAT 403 454 FG-GAP 4.

```
FT REPEAT 456 517 FG-CAP 5.
FT REPEAT 519 577 FG-CAP 6.
FT REPEAT 582 634 FG-CAP 7.
FT CA BIND 467 475 Potential.
FT CA BIND 531 539 Potential.
FT CA BIND 594 602 Potential.
FT SITE 1126 1130 GPKR motif.
FT DISULFID 69 76 By similarity.
FT DISULFID 108 126 By similarity.
FT DISULFID 656 711 By similarity.
FT DISULFID 769 775 By similarity.
FT DISULFID 845 860 By similarity.
FT DISULFID 993 1017 By similarity.
FT DISULFID 1022 1027 By similarity.
FT CARBOHYD 61 61 N-linked (GlcNAc...) (potential).
FT CARBOHYD 89 89 N-linked (GlcNAc...) (potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (potential).
FT CARBOHYD 393 393 N-linked (GlcNAc...) (potential).
FT CARBOHYD 696 696 N-linked (GlcNAc...) (potential).
FT CARBOHYD 734 734 N-linked (GlcNAc...) (potential).
FT CARBOHYD 784 784 N-linked (GlcNAc...) (potential).
FT CARBOHYD 907 907 N-linked (GlcNAc...) (potential).
FT CARBOHYD 936 936 N-linked (GlcNAc...) (potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc...) (potential).
SQ SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;

Query Match 18.6%; Score 1136; DB 1; Length 1161;
Best Local Similarity 30.1%; Pred. No. 9e-72;
Matches 363; Conservative 208; Mismatches 486; Indels 148; Gaps 46;

Qy 20 CSPNLDEHRLPPGPEAFGVSVLQHVGGGQRMVLVGAPEWDGPGSDRGDGVYRCPVG 79
Db 17 CHGSLNDEEPIVF-REDAASFGTVVQF--GSGR-LVVGAPLEAVAVNQTRLYDC--- 69
Qy 80 GAHNAPCAKHLGQYQLGNSHPVNMHLGMSLLETDCDGGFWACAPLWSRACSSVFS 139
Db 70 ----AP-ATGWCQPIVL-RSPLEAVNMSLSLVTATNNAQLLACGTPAQACVKNMYAK 123
Qy 140 GICARVDASFPQGSGLAPTAQRCTY--MDVVIVLDGNSI--YPMSEVQTFRLRLVKLFP 196
Db 124 GSCLLGSSLFQIQAVPASMCEPCREQEMDIAFLIDGSGSINQORDFAQMKDFVKALMGE-F 182
Qy 197 IDPQIQVGLVOYGESPVHESLGDERTKEVRVRAKNLSRREGRETKTAQAINVACTEG 256
Db 183 ASTSTL-FSLMQYSNLIKTHFTTFEKNILDQSLVDPIVQLGL-TYATGIRTVMEEL 240
Qy 257 FSQSHGGRPEARLLVVVTDGESH-DGEELPAALKAACEAGRVTRYGIAVLGHYLRQRDP 315
Db 241 FHSKNGSRKSAKKILLVITDQKYRDPLEYSDVIPADKAGIIRYAICVGDAF---QEP 296
Qy 316 SSFLREIRTIASDPDRFFFNVTDEAALTDIVDALGDRIFGLEGSHAENESSFGLEMSQI 375
Db 297 TA-LKELNTIGSAPPQDHPVFKVGNFAALRSIQRLQEKIFAIEGTQSRSSSSFOHEMSQE 355
Qy 376 GFSTHRLKDGTLFGMVGAYDGGSVLWLEGHRLPPP-----RMALEDEFPALQNH 428
Db 356 GFSSALTSDBGVLGAVGSFSW-----SGGAFLYPPNTRFTFINMQEN-----VDMRD 403
Qy 429 AYLGYSSVSMLLRGRRLLFSGCAPFRHRGVIAFOLKKGAVRVAQSLQGEQIGSYFGS 488
Db 404 SYLGYSTAVAFWKGVHSLIL-GARPHQHTGVVIF-TQEARHWRPKSEVRGTQIGSYFGA 461
Qy 489 ELCPLDTRDGTDTDLVLAAPMFLGPQNKETGRVYVILV-CQQSLLLTLQGTLOPEPPQD- 546
Db 462 SLCSVDVDRDGSITDLVLIGAPHYY--EQTRGGQSVFPVPGVGRGWQCEATLHGEQGHWP 519
Qy 547 ARFGFAMGALPDNLQDGFADVAVGAPLEDHGOALYLYHG-TQSVRPHPAQRIAAASMP 605
Db 520 GRFGVALTVLGDVNGDNLADVAIGAPGEESRGAVYIFHGASRLIEIMPSPQRVTGSQLS 579
Qy 606 HALSYFGSRVDRLDLDGDDLDVAVGAQGAAILLSRRPIVHLTPSLVTPQAISVWQRD 665
Db 580 LRLQYFGSLSGGQDLTQDGLVDLAVGAQGHVLLSLPLKVELSIFAPMEVAKAVYQ 639
```

```
Qy 666 CRRR-----QGEAVCLTAALCFQVYTSRTPGRWDH-QFYMRFTASLDEWTAGARAADFG 717
Db 640 CWERTPTVLEAGEATVCLT-----VHKGSPDLLGNVQGSVRYVDLALDPGLRLISRAIFDE 693
Qy 718 SQQLSPRRRLRSVGNVTCBQLHFPVLVD-TSDYLRPVALVTFTAL-DNTTKP---GPVLN 772
Db 694 TKNCTLTGRKTLGLD-HCETVKLLLPDCVEDAVSPILRLNLSLVDSDASPRNLHPVLA 752
Qy 773 EGSPTSIOQLVPFSKDCGPNCEVTDLVLQNMDIRSKAPFVVRGGRKRVLVSTTLEN 832
Db 753 VGSQDHITASLPFEKNCKQELLCEGLGISFNFS-----GLQVLVVGSGSPELTVTVTVN 807
Qy 833 RKNAYNTSLSIIFSRNLHLASLTPQRES---PIKVEC-RAPSAAHRL-----CSVGHVPV 884
Db 808 EGSDSYGLTVKFTYPAGLSVRRVTGTQPHQYPLRLACEAPAAQEDLRSSCSINHPIF 867
Qy 885 QTGAQVTFLEFBFSCSSLSQVFGKLTASSDSLSERNGLTQENTA-QTSAYIQY----- 937
Db 868 REGAKTTFMITFDVSYKAFGLDRL--LLRAKASSENKPDNTKTAFOLELPVKYTVYTLI 925
Qy 938 -----EPHLLFSS-----ESTLHRYEVHPYGTLPVGPGEPEFTTLRVQNLGCVVSG 985
Db 926 SRQEDSTNHVNFSSSHGGRQEAHRYRVNLSPL-----KLAVRVNFWFVPLLVNGV 977
Qy 986 IISALLPAVAGGNYFSLSQVITNNASCIVQNLTEPPGPPVHPPEELOHTNRLNGSNTQ 1044
Db 978 AVMDVTLLSSPAQG-----VSCVSO--MKPPONPDPFLTQIORSVLDCSIAD 1021
Qy 1045 QQVVRCHLGQAKGTEVSVGLLRLVHNEFFRRRAKFKSLTVVSTPELGTGEGSVLQL--TE 1102
Db 1022 CLHFRCDIPSLDIQDELDFILRGNLSFGWVSQTLQEKVLLVSEAEITFDTSVYSQLPQOE 1081
Qy 1103 A---SRWSESLLEVQTRPILISLWILIGSVLGGLLLLALLVFCWLKGLGFFPAHKKIPBE 1159
Db 1082 AFLRAQVETTLEEVVVEPI----FLVAGSVGGLLLALLITVVLVKLGFF----- 1128
Qy 1160 KREEK 1164
Db 1129 KRQYK 1133
```

Search completed: April 6, 2005, 12:15:23

Job time : 179.72 secs

This Page Blank (uspto)

DT	26-AUG-2004	(first entry)	
DE	Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.		
XX	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.		
XX	Homo sapiens.		
XX	WO2004048938-A2.		
XX	10-JUN-2004.		
XX	26-NOV-2003; 2003WO-US038193.		
XX	26-NOV-2002; 2002US-0429739P.		
XX	(PROT-) PROTEIN DESIGN LABS INC.		
XX	Aziz N, Ginsburg WM, Zlotnik A;		
XX	WPI; 2004-441208/41.		
XX	Early detection of soft tissue sarcoma comprises determining expression		
XX	of a gene in a first soft tissue sample and a normal soft tissue sample		
XX	and comparing the gene expression, also useful in treating soft tissue		
XX	sarcoma.		
PS	Example 2; SEQ ID NO 2109; 210pp; English.		
CC	The invention relates to a novel method for detecting soft tissue sarcoma		
CC	which comprises obtaining a first soft tissue sample from an individual		
CC	and a normal soft tissue sample from the same or different individual,		
CC	determining the expression of a gene in both samples and comparing the		
CC	expression of the gene in both soft tissue samples, where a higher level		
CC	of protein expression in the first soft tissue sample indicates the		
CC	presence of soft tissue sarcoma. The method of the invention has		
CC	cytostatic applications and may be useful for detecting soft tissue		
CC	sarcoma, possibly via gene therapy or vaccine production. The nucleic		
CC	acid sequences may be useful in diagnostic and screening applications.		
CC	The current sequence is that of a human soft tissue sarcoma-upregulated		
CC	protein of the invention. The current sequence is not shown within the		
CC	specification per se but was submitted in CD format by the inventor.		
XX	Sequence 1167 AA;		
XX	Query Match 99.8%; Score 6092; DB 8; Length 1167;		
XX	Best Local Similarity 99.7%; Pred. No. 0;		
XX	Matches 1164; Conservative 2; Mismatches 1; Indels 0; Gaps 0;		
Qy	1 MELPFTVTHLFLPLVFLTGLCSPPFNLDHHPRLFPGPPEAFGYSVLQHVGGQRWMLVGA 60		
Db	1 MELPFTVTHLFLPLVFLTGLCSPPFNLDHHPRLFPGPPEAFGYSVLQHVGGQRWMLVGA 60		
Qy	61 PWDGSGDRGRGVRCVPGGAHNPACAKGHLGDYQLGNSSHPAVNMHLGMSLLTDDGG 120		
Db	61 PWDGSGDRGRGVRCVPGGAHNPACAKGHLGDYQLGNSSHPAVNMHLGMSLLTDDGG 120		
Qy	121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTYMDVIVLDGNSNIYP 180		
Db	121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTYMDVIVLDGNSNIYP 180		
Qy	181 WSEVQTFLLRLVGLKLFIDPEIQVGLVQYGSPPVHWSLGDPRFRTKEEVVRAAKNLSREG 240		
Db	181 WSEVQTFLLRLVGLKLFIDPEIQVGLVQYGSPPVHWSLGDPRFRTKEEVVRAAKNLSREG 240		
Qy	241 RETKTAQIAWVACTEGFSQSGHGRPEAARLLVWTDGSHDGEELPAALKAACEAGRVTY 300		
Db	241 RETKTAQIAWVACTEGFSQSGHGRPEAARLLVWTDGSHDGEELPAALKAACEAGRVTY 300		
Qy	301 GIAVLGHLVRRQDPSSFLRIRTIATSDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360		
Db	301 GIAVLGHLVRRQDPSSFLRIRTIATSDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360		

Qy	361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDWGCVLWLEGHRLPPRMALEDEF 420		
Db	361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDWGCVLWLEGHRLPPRMALEDEF 420		
Qy	421 PPALQNHAAVLYGYSVSSMLLRGGRRFLSGAPRFRHKGKVIAFOLKDKGAVRVAQSLOGE 480		
Db	421 PPALQNHAAVLYGYSVSSMLLRGGRRFLSGAPRFRHKGKVIAFOLKDKGAVRVAQSLOGE 480		
Qy	481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTLLQ 540		
Db	481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTLLQ 540		
Qy	541 PEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHOGALYLVHGTQSGVRPHPAQRIA 600		
Db	541 PEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHOGALYLVHGTQSGVRPHPAQRIA 600		
Qy	601 AASMPHALSYFGRSVDGRLDLDDGDDLDVAVAGAACAAIILLSSRPVHLTSPLEVTPOAIS 660		
Db	601 AASMPHALSYFGRSVDGRLDLDDGDDLDVAVAGAACAAIILLSSRPVHLTSPLEVTPOAIS 660		
Qy	661 VVQDCCRGGOEAVCLTAALCFQVTSRTPGRWDHGFYMRFTASLDEWTAGARAAFDGSGQ 720		
Db	661 VVQDCCRGGOEAVCLTAALCFQVTSRTPGRWDHGFYMRFTASLDEWTAGARAAFDGSGQ 720		
Qy	721 RLSRRLRLSVGNVTCBQLHFLVLDTSYLRPVALTVTFALDNTTKGPNVLEGSPTSIO 780		
Db	721 RLSRRLRLSVGNVTCBQLHFLVLDTSYLRPVALTVTFALDNTTKGPNVLEGSPTSIO 780		
Qy	781 KLVPFSKDCPDNECVTDVLQVNMDIRGSKAPFVVRGGRKVLVSTTLNKRKENAYNT 840		
Db	781 KLVPFSKDCPDNECVTDVLQVNMDIRGSKAPFVVRGGRKVLVSTTLNKRKENAYNT 840		
Qy	841 SLSIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPFQTGAKVTFLLFEFESC 900		
Db	841 SLSIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPFQTGAKVTFLLFEFESC 900		
Qy	901 SLLSQVFGKLTASSDSLRNGTLQENTQTAQTSAYIOEPHLLFSESTLHREYVHPYGT 960		
Db	901 SLLSQVFGKLTASSDSLRNGTLQENTQTAQTSAYIOEPHLLFSESTLHREYVHPYGT 960		
Qy	961 PVGPGPFKTTLRVQNLGCVVSGLIISALLPAVAGHGNFSLSQVITNNASCIVQNL 1020		
Db	961 PVGPGPFKTTLRVQNLGCVVSGLIISALLPAVAGHGNFSLSQVITNNASCIVQNL 1020		
Qy	1021 EPPGPPVHPELQHTNRLNGSNTCCQVVRCHLGQAKGTEVSVGLLRVHNEFFRRAKFK 1080		
Db	1021 EPPGPPVHPELQHTNRLNGSNTCCQVVRCHLGQAKGTEVSVGLLRVHNEFFRRAKFK 1080		
Qy	1081 SLTVVSTFELGTEGSLVQLTEASRWSESLLEVVQTRPILISLWILGSLGULLLALL 1140		
Db	1081 SLTVVSTFELGTEGSLVQLTEASRWSESLLEVVQTRPILISLWILGSLGULLLALL 1140		
Qy	1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167		
Db	1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167		
XX	RESULT 3		
XX	AAB64584		
ID	AAB64584 standard; protein; 1167 AA.		
XX	AC AAB64584;		
XX	DT 22-MAR-2001 (first entry)		
XX	DE Human secreted protein #37.		
XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;		
XX	antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;		
XX	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;		
XX	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;		
XX	neurological disease; infection; human; secreted protein.		

XX OS Homo sapiens.
XX PN WO2003000865-A2.
XX PD 03-JAN-2003.
XX PF 26-MAR-2002; 2002WO-US009105.
XX PR 27-MAR-2001; 2001US-0278650P.
XX PR 12-SEP-2001; 2001US-00950082.
XX PR 12-SEP-2001; 2001US-00950083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR N-PSDB; ADA43832.
XX PT A human secreted protein and nucleic acids useful for preparing a
PT diagnostic or pharmaceutical composition for diagnosing or treating
PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,
PT retinopathy, neuropathy.
XX PS Claim 1; SEQ ID NO 214; 701pp; English.
XX CC The invention relates to novel genes and their fragments which are useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids and proteins
CC are useful in the diagnosis, treatment and prevention of conditions
CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,
CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,
CC infection, cataract, renal disorders, or endocrine disorders. The present
CC sequence was used to illustrate the invention.
XX SQ Sequence 1167 AA;
Query Match 98.9%; Score 6040; DB 6; Length 1167;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1155; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
Qy 1 MELPFVTHLFLPLVFLTGLCSPPFNLDHHPRLFPGPPEAEFGYSLQHVGGQRMVLVGA 60
Db :
1 VELPFVTHLFLPLVFLTGLCSPPFNLDHHPRLFPGPPEAEFGYSLQHVGGQRMVLVGA 60
Qy 61 PWDGSGDRRGDVYRCVGGAHNAPCAKGLHLDYQLGNSHPVNMHLGMSLLETGDGG 120
Db :
61 PWDGSGDRRGDVYRCVGGAHNAPCAKGLHLDYQLGNSHPVNMHLGMSLLETGDGG 120
Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVIVLDGNSNIYP 180
Db :
121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVIVLDGNSNIYP 180
Qy 181 WSEVOTFLRLRVGLKFLIDPEIQVGLVQYQVGSPPVHWSLGDFTKKEVVRRAKNLSRREG 240
Db :
181 WSEVOTFLRLRVGLKFLIDPEIQVGLVQYQVGSPPVHWSLGDFTKKEVVRRAKNLSRREG 240
Qy 241 RETYTAQAIWVACTEGSQSHGGRPEARLLVVTGESHDPGEELPAALKAACEAGRVTRY 300
Db :
241 RETYTAQAIWVACTEGSQSHGGRPEARLLVVTGESHDPGEELPAALKAACEAGRVTRY 300
Qy 301 GIAVLGHLRRQRPSSFLRIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360
Db :
301 GIAVLGHLRRQRPSSFLRIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKDXGLFGMVGAYDWGGSVLWLEGCHRLFPFRMALEDEF 420
Db :
361 HAENESSFGLMSQIGFSTHRLKDXGLFGMVGAYDWGGSVLWLEGCHRLFPFRMALEDEF 420
Qy 421 PPALQNHAXYLGYSSXWMLRGXRLXLSGAXRFRHRGKVIAFOLKKGAVVRAQSLQGE 480

Db 421 PPALQNHAXYLGYSSXWMLRGXRLXLSGAXRFRHRGKVIAFOLKKGAVVRAQSLQGE 480
Qy 481 QIGSYFGSELCPDLTDRTDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQQLLTQGTFLQ 540
Db :
481 QIGSYFGSELCPDLTDRTDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQQLLTQGTFLQ 540
Qy 541 PEPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGHQGALYLHGTQSGVRPHPAQRIA 600
Db :
541 PEPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGHQGALYLHGTQSGVRPHPAQRIA 600
Qy 601 AASMPHALSYFGRSVDGRDLDDGDDLDVAVGACAAAILLSSRPVHLTSPLEVTPOAIS 660
Db :
601 AASMPHALSYFGRSVDGRDLDDGDDLDVAVGACAAAILLSSRPVHLTSPLEVTPOAIS 660
Qy 661 VVQDCRRRGOEAVCLTAALCFQVTSRTPGWDHOFYMRFTASLDEWTAGARAAFDGSGQ 720
Db :
661 VVQDCRRRGOEAVCLTAALCFQVTSRTPGWDHOFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSRRLRLSVGNVTCQLHFPVLDLSDYLRPVALTVTFALDNTTKGPPVINEGSPISIQ 780
Db :
721 RLSRRLRLSVGNVTCQLHFPVLDLSDYLRPVALTVTFALDNTTKGPPVINEGSPISIQ 780
Qy 781 KLVPFSDKCGPDNECVTDLVQNMDIRGSKAPFVVRGGRKVLVSTTLNRRKENAYNT 840
Db :
781 KLVPFSDKCGPDNECVTDLVQNMDIRGSKAPFVVRGGRKVLVSTTLNRRKENAYNT 840
Qy 841 SLSIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFOTGAKVTFLEFEFSC 900
Db :
841 SLSIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFOTGAKVTFLEFEFSC 900
Qy 901 SLLSQVFGKLTASSDSLERNGTLOENTAQTSAYIOYEPHLLFSSESTLHRYEYHPYCTL 960
Db :
901 SLLSQVFGKLTASSDSLERNGTLOENTAQTSAYIOYEPHLLFSSESTLHRYEYHPYCTL 960
Qy 961 PVGPGPEFKTTLRVQNLGCVVSGLIISALLPAVAGHGNFELSLSQVTTNNASCIVQNL 1020
Db :
961 PVGPGPEFKTTLRVQNLGCVVSGLIISALLPAVAGHGNFELSLSQVTTNNASCIVQNL 1020
Qy 1021 EPPGPPVHPEELQHTNRLNGSNTQCQVVRCHLQAGTQVSVGLLRVHNEFFRRAKFK 1080
Db :
1021 EPPGPPVHPEELQHTNRLNGSNTQCQVVRCHLQAGTQVSVGLLRVHNEFFRRAKFK 1080
Qy 1081 SLTVVSTFELGTEGSLVQLTEASRWSESLLEVQVTRPILISLWILGSLGGILLALL 1140
Db :
1081 SLTVVSTFELGTEGSLVQLTEASRWSESLLEVQVTRPILISLWILGSLGGILLALL 1140
Qy 1141 VFCLWKLGFPAHKKIPPEEKREKLEQ 1167
Db :
1141 VFCLWKLGFPAHKKIPPEEKREKLEQ 1167
RESULT 7
AAB64657
ID AAB64657 standard; protein; 1152 AA.
XX AAB64657;
AC AAB64657;
DT 22-MAR-2001 (first entry)
XX Human secreted protein BLAST search protein SEQ ID NO: 167.
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX Homo sapiens.
XX OS
XX PN WO200077197-A1.
XX PD 21-DEC-2000.

(HUMA-)	HUMAN GENOME SCI INC.
(ROSE/)	ROSEN C A.
Rosen CA,	Ruben SM, Komatsoulis GA;
WPI;	2001-032312/04.
Isolated	nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition.
Disclosure;	Page 547-551; 558pp; English.
The invention	relates to the isolation of genes AAP32757-F32803 encoding the human secreted proteins AAB64549-B64594. The sequence is used as a query sequence for doing BLAST searches to identify homologous sequences. The genes and proteins are useful for preventing, ameliorating or treating medical conditions, e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies; autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections
Query Match	97.7%; Score 5965; DB 4; Length 1152;
Best Local Similarity	99.0%; Pred. No. 0;
Matches 1140;	Conservative 3; Mismatches 9; Indels 0; Gaps 0;
Qy	1 MELPFVTHLPPLVFVLGLCPNLDEHPRLFPGPPEAEFGYSVLQHVGSGQRMLYGA 60
Db	: 1 VELPFVTHLPPLVFVLGLCPNLDEHPRLFPGPPEAEFGYSVLQHVGSGQRMLYGA 60
Qy	61 PWDGPGDRRGDYYRCPVGGAHNAPCAKHLGDYQLGNSSHPAVNMHLGMSLLETDDGG 120
Db	61 PWDGPGDRRGDYYRCPVGGAHNAPCAKHLGDYQLGNSSHPAVNMHLGMSLLETDDGG 120
Qy	121 FMACAPLWSRACGSVPFSSGICARVDASFOQPSGLAPTAAQCRTPTMYDVIVLDGNSIYP 180
Db	121 FMACAPLWSRACGSVPFSSGICARVDASFOQPSGLAPTAAQCRTPTMYDVIVLDGNSIYP 180
Qy	181 WSEVQTFLRLVKLFIDPEIQIGLVQYGSPVHEWSLGDFRTKEEVRAAKNLSREG 240
Db	181 WSEVQTFLRLVKLFIDPEIQIGLVQYGSPVHEWSLGDFRTKEEVRAAKNLSREG 240
Qy	241 RETKTQAQIMVACTEFSQSNGHRPBAERLLVVVTDGESHGGEELPAALKACEAGRVTRY 300
Db	241 RETKTQAQIMVACTEFSQSNGHRPBAERLLVVVTDGESHGGEELPAALKACEAGRVTRY 300
Qy	301 GIAVLGHYLRRORDPSSELREIRTIASDPDERFFNVTDDEAALTDIVDALGRIFGLEGS 360
Db	301 GIAVLGHYLRRORDPSSELREIRTIASDPDERFFNVTDDEAALTDIVDALGRIFGLEGS 360
Qy	361 HAENESSFGLMSEQISGFSTRHLKDGLIFGMVAYDWGGSVMLEGGHRLFPPRMALDEF 420
Db	361 HAENESSFGLMSEQISGFSTRHLKDGLIFGMVAYDWGGSVMLEGGHRLFPPRMALDEF 420
Qy	421 PPALQNHAAYLYGVSVSMILLRGRRILFLSAPFRHRGKVIAFOLKDCGAVRVAOSLOQE 480
Db	421 PPALQNHAAXLYGVSVSMILLRGXRKLKSAXRFRRHRGKVIAFOLKDCGAVRVAOSLOQE 480
Qy	481 QIGSYGSELCLDTDRDGTDLVLAAMPFLGPONKETGRVYVYLVGOOSLTTLTQGTLQ 540
Db	481 QIGSYGSELCLDTDRDGTDLVLAAMPFLGPONKETGRVYVYLVGOOSLTTLTQGTLQ 540
Qy	541 PEPPQDAFCFAMGALPDNLNDGCFADVAVGAPLEDHGCHOGALYLEHTGTSGVRPHFAORIA 600

```

PR 02-APR-1998; 98SE-00001164.
PR 28-JAN-1999; 99SE-00000319.
XX (ACT1-) ACTIVE BIOTECH AB.
XX
XX Lundgren-Akerlund E;
XX
XX WPI; 2000-052639/04.
XX N-PSDB; AAZ34720.
XX
XX New isolated integrin subunit alpha-10, used as a marker or target
XX molecule for cells during development, regeneration and pathological
XX conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
XX inflammation.
XX
XX Claim 1; Page 43-48; 90pp; English.
XX
XX This sequence represents a splice variant of novel human chondrocyte
XX integrin subunit alpha-10 (I8a10). It is identical to I8a10 (see
XX CC AAY3242) except for deletion of amino acids 975-986. The invention
XX relates to a recombinant or isolated integrin heterodimer comprising the
XX alpha10 subunit in association with subunit beta (especially beta-1). The
XX heterodimer, subunit alpha-10 or splice variant can be used as a marker
XX or target of all types of cells, e.g. of chondrocytes, osteoblasts and
XX fibroblasts. They can also be used: for treating pathological conditions
XX involving I8a10, such as damage to cartilage, trauma, rheumatoid
XX arthritis or osteoarthritis; for detecting the formation of cartilage
XX during embryonal development, physiological or therapeutic repair of
XX cartilage, or detecting regeneration of cartilage or chondrocytes during
XX transplantation of cartilage or chondrocytes; for selection and analysis
XX or for sorting, isolating or purification of chondrocytes and for in
XX vitro studies of differentiation of chondrocytes; and as a target for
XX anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or
XX other tissues where adhesion impairs the function of the tissue (all
XX claimed). I8a10 binding entities can be used to determine the
XX differentiation-state of cells during embryonic development, angiogenesis
XX or development of cancer, in pathological conditions such as rheumatoid
XX arthritis, osteoarthritis or cancer, in tissue regeneration or in
XX therapeutic and physiological repair of cartilage (claimed). A
XX vaccine comprising the integrin heterodimer or subunit alpha-10 is also
XX claimed. I8a10 polynucleotides, vectors, host cells and methods of
XX producing recombinant I8a10 are also claimed
XX
XX Sequence 1132 AA;
XX
XX Query Match 96.7%; Score 5904.5; DB 3; Length 1132;
XX Best Local Similarity 97.0%; Pred. No. 0;
XX Matches 1132; Conservative 0; Mismatches 0; Indels 35; Gaps 1;
XX
QY 1 MELPFVTHLFLPLVFLTGLCSPPFNLDHHPRLFPGPPEAFGYSVLQHVGGQRWMLVGA 60
DB 1 MELPFVTHLFLPLVFLTGLCSPPFNLDHHPRLFPGPPEAFGYSVLQHVGGQRWMLVGA 60
XX
QY 61 PWDGSGDRGDVYRCPVGGAAHAPCAKGLHDYQLGNSHPA VNMHLGMSLLETDDGG 120
DB 61 PWDGSGDRGDVYRCPVGGAAHAPCAKGLHDYQLGNSHPA VNMHLGMSLLETDDGG 120
XX
QY 121 FMACAPLWSRACGSSVSSGICARVDASFPQGS LAPTAQRCPYMDVIVLDGNSIYP 180
DB 121 FMACAPLWSRACGSSVSSGICARVDASFPQGS LAPTAQRCPYMDVIVLDGNSIYP 180
XX
QY 181 WSEVQTFLLRLVGLKFLDPEIQVGLVQGESPVHWSLGDFTKKEEVVRAAKNLSRREG 240
DB 181 WSEVQTFLLRLVGLKFLDPEIQVGLVQGESPVHWSLGDFTKKEEVVRAAKNLSRREG 240
XX
QY 241 RETKTAQIMVACTEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRY 300
DB 241 RETKTAQIMVACTEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRY 300
XX
QY 301 GIAVLGHVLRQRDPSSFLRIRTIASDPDRFFPNVTDEAALTDIVDALGDRIFGLEGS 360
DB 301 GIAVLGHVLRQRDPSSFLRIRTIASDPDRFFPNVTDEAALTDIVDALGDRIFGLEGS 360
XX

```

RESULT 10

ADR41424

ID ADR41424 standard; protein; 1049 AA.

XX

XX ADR41424;

XX AC

XX 07-OCT-2004 (first entry)

XX

XX Human CD-like molecule HAOAD02, SEQ ID NO:223.

XX

XX Human; CD-like molecule; cluster of differentiation; diagnosis;

XX prevention; immune disorder; immunodeficiency; autoimmune disorder;

XX blood-related disorder; haematological disorder; haemostatic disorder;

XX thrombolytic disorder; hyperproliferative disorder; cancer; tumour;

XX apoptotic disorder; cardiovascular disorder; respiratory disorder;

XX angiogenic disorder; neovascularisation; neurological disorder;

XX

endocrine disorder; reproductive system disorder; infectious disease; gastrointestinal disorder; drug screening; tissue regeneration; chemotaxis; gene therapy; antibody therapy; drug targeting; chromosome mapping; forensic analysis; immunophenotyping; cytostatic; haemostatic; tranquiliser; vulnery; antiinflammatory; nephrotropic; cardiant; antiallergic; anti-HIV; antirheumatic; antiarthritic; antipruritic; immunosuppressive; vasotrophic; neurotropic; antithyroid; thymimetic; gynaecological; virucide; hepatotropic; antibacterial; dermatological; chromosome 1q21.

Homo sapiens.
 WO200226930-A2.
 04-APR-2002.
 25-SEP-2001; 2001WO-US029838.
 26-SEP-2000; 2000US-0235484P.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Birse CE;
 WPI; 2002-405050/43.
 N-PSDB; ADR41248.
 Novel polynucleotides and polypeptides useful for treating, preventing or ameliorating cardiovascular, renal, neurovascular, and autoimmune disorders.
 Claim 11; SEQ ID NO 223; 1243pp; English.
 The invention relates to 167 novel human CD (cluster of differentiation) - like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid:11)-
 Sequence 1049 AA;

Query Match 83.2%; Score 5080; DB 5; Length 1049;
 Best Local Similarity 94.5%; Pred. No. 0;
 Matches 990; Conservative 7; Mismatches 31; Indels 20; Gaps 3;

Qy 124 CAPLWSRA-CGSSVFSGICARVDASFOQSLAPTAQRCPTYMDV---VIVLGSNSIY 179
 Db 18 CLQFWMDCPCGCFIPASG-----KPGTHCPTLPNIHGCHCLGLHSIY 61
 Qy 180 PWEVQTFLLRLVGLKFLTDPEQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRE 239
 Db 62 PWEVQTFLLRLVGLKFLTDPEQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRE 121
 Qy 240 GRETKTAQAIMVACTEGFSQSHGRPEARLLVVVTDGESHGDBELPAALKACEAGRVTR 299
 Db 122 GRETKTAQAIMVACTEGFSQSHGRPEARLLVVVTDGESHGDBELPAALKACEAGRVTR 181
 Qy 300 YGIAVLGHLRRQDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEG 359
 Db 182 YGIAVLGHLRRQDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEG 241
 Qy 360 SHAENESSFGLMSQIGFSTHRLKDGILFGVMGAYDWGSGVLWLEGGHRLFPFPAWLEDE 419
 Db 242 SHAENESSFGLMSQIGFSTHRLKDGILFGVMGAYDWGSGVLWLEGGHRLFPFPAWLEDE 301
 Qy 420 FPPALQNHAAVYLGYSVSSMLRGRRFLSGAPRFRHRGKVIAFOLKKGAVRVAQSLQG 479
 Db 302 FPPALQNHAAVYLGYSVSSMLRGRRFLSGAPRFRHRGKVIAFOLKKGAVRVAQSLQG 361
 Qy 480 EQIGSYFGSELCPDLDTRDGTDLVLLVAAAPFLGPNKQKTRGVYVYLVGQSSLLTLQGT 539
 Db 362 EQIGSYFGSELCPDLDTRDGTDLVLLVAAAPFLGPNKQKTRGVYVYLVGQSSLLTLQGT 421
 Qy 540 QPEPQDARFGFAMGALPDNODGFADVAVGAPLEDHGGALYLYHGTQSGVRHPAORI 599
 Db 422 QPEPQDARFGFAMGALPDNODGFADVAVGAPLEDHGGALYLYHGTQSGVRHPAORI 481

Qy 600 AAASMPHALSYFGRSVDRGLDLDGDDLVTVAVGAQGAAILSSRPVHLTPSLEVTPOAI 659
 Db 482 AAASMPHALSYFGRSVDRGLDLDGDDLVTVAVGAQGAAILSSRPVHLTPSLEVTPOAI 541
 Qy 660 SVVQDRCRRGQEAVALCTAALCFQVTSRTQGRWDHQFMRFTASLDWTAGARAAFDGSG 719
 Db 542 SVVQDRCRRGQEAVALCTAALCFQVTSRTQGRWDHQFMRFTASLDWTAGARAAFDGSG 601
 Qy 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTTFALDNTTKPGPVINEGSPTSI 779
 Db 602 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTTFALDNTTKPGPVINEGSPTSI 661
 Qy 780 QKLVPPSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGRRKVLVSTTLENKENAYN 839
 Db 662 QKLVPPSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGRRKVLVSTTLENKENAYN 721
 Qy 840 TSLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFQTAGKVTFLLEPEFS 899
 Db 722 TSLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFQTAGKVTFLLEPEFS 781
 Qy 900 CSSLLSQVFGKLTASSDSLERNGTLQENTAQTSAYIQVEPHLLFSSESTLHRYEVHPYGT 959
 Db 782 CSSLLSQVFGKLTASSDSLERNGTLQENTAQTSAYIQVEPHLLFSSESTLHRYEVHPYGT 841
 Qy 960 LPVGPGEFETTLRVQNLGCVVSGLLISALLPAVAGGNYFLSLQVITNASCIVQNL 1019
 Db 842 LPVGPGEFETTLRVQNLGCVVSGLLISALLPAVAGGNYFLSLQVITNASCIVQNL 901
 Qy 1020 TEPGPPVHPEELQHTNRLNGSNQCCVVRCHLGQAKGTEVSGLLRVLVNEPFRRAKF 1079
 Db 902 TEPGPPVHPEELQHTNRLNGSNQCCVVRCHLGQAKGTEVSGLLRVLVNEPFRRAKF 961
 Qy 1080 KSLTVVSTFELGTBEGSVLQTEASRWSESLLEVQTRPILISLWILIGSVLGGLLLAL 1139
 Db 962 KSLTVVSTFELGTBEGSVLQTEASRWSESLLEVQTRPILISLWILIGSVLGGLLLAL 1021
 Qy 1140 LVFLCWKLGFPAHKKIPEEEKREKLEQ 1167
 Db 1022 LVFLCWKLGFPAHKKIPEEEKREKLEQ 1049

RESULT 11

AAU14231
 ID AAU14231 standard; protein; 1188 AA.

AC AAU14231;

DT 24-OCT-2001 (first entry)

XX Human novel protein #102.

Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
 immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
 anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 antibacterial; antiallergic; dermatological; haemostatic; antidiabetic;
 thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 tissue regeneration; immune disorder.

XX Homo sapiens.

XX WO200155437-A2.

PD 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US002623.

XX 25-JAN-2000; 2000US-00491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

Human A259 polypeptide.

Human, A259; integrin alpha subunit; integrin alpha 10; secreted protein; liver disease; fibrosis; lung; kidney; bone associated disorder; blood; cartilage associated disorder; hematopoietic disorder; bone marrow; immune related disease; apoptotic disorder; neuronal tissue disease; neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic; nephrotropic; immunomodulator; anti-inflammatory; neuroprotective; antiarthritic; antianemic; antiallergic; antiaesthetic; dermatological; antidiabetic; anticonvulsant; antiparkinsonian.

Homo sapiens.

Key Location/Qualifiers
Domain 1..1141
/note= "Extracellular domain"
Peptide 1..22
/note= "Signal peptide"
Protein 23..1188
/note= "Mature human A259"
Domain 37..90
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
Domain 115..157
/note= "Integrin alpha repeat domain"
Domain 164..345
/note= "I domain or Von Willebrand Factor type A domain"
Domain 367..392
/note= "Integrin alpha repeat domain"
Domain 421..472
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
Domain 476..532
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
Domain 538..593
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
Domain 600..654
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
Domain 1142..1164
/note= "Transmembrane domain"
Domain 1165..1188
/note= "Cytoplasmic domain"

WO200181414-A2.

01-NOV-2001.

27-APR-2001; 2001WO-US013516.

27-APR-2000; 2000US-00561263.

(MILL-) MILLENNIUM PHARM INC.

Pan Y, Lora J;

WPI; 2002-041397/05.

N-PSDB; AAS16873.

New A259 nucleic acids and polypeptides, which comprise integrin alpha subunit, useful for diagnosing, preventing or treating e.g. liver disease, kidney or lung fibrosis, cancers, blood disorders or immune related diseases.

Claim 9; Fig 1; 168pp; English.

The invention relates to human and murine A259 nucleic acid molecules which encode secreted proteins with homology to integrin alpha subunits, specifically to integrin alpha 10. The A259 polypeptide and nucleic acid are useful for treating liver disease or fibrosis, particularly kidney fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also

CC useful for diagnosing, preventing or treating cartilage and bone associated disorders (such as bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and osteoporosis), bone marrow, blood and hematopoietic disorders (such as acute myeloid leukemia, haemophilia, anaemia and thalassemia), immune related diseases (such as HIV, viral infections, cancers, T cell autoimmunity disorders e.g. AIDS, and allergic inflammatory disorders e.g. asthma and psoriasis), apoptotic disorders (such as systemic lupus erythematosus and insulin-dependent diabetes mellitus), diseases of the neuronal tissues (such as epilepsy and muscular dystrophy) and neurodegenerative diseases (such as Parkinson's disease and Huntington's disease). This sequence represents the human A259 polypeptide

XX SQ Sequence 1188 AA;

Query Match 40.0%; Score 2439.5; DB 5; Length 1188;

Best Local Similarity 43.2%; Pred. No. 8.5e-208; Indels 45; Gaps 16;

Matches 517; Conservative 213; Mismatches 421;

Qy 1 MELPFVTHLEPLVFLTGLCSFNLDSHHPRLPGPPPEARFYSVLQHVGGQRMVLVGA 60
Db 1 MDLPRLGLVVAWALSLLWPGFTDTFNMTRKPRVPGSRTAFPGYTVQOHDISGNKMLVGA 60
Qy 61 PMDPSGDRRGDYVRCVPGGAHNAPCAKHGLDYQLGNSHHPVNMHLGMSLLETDDGG 120
Db 61 PLETNGYQKTGDVYKCPV---IHGNCCTKMLGRVTLNSVSEKDNMRLGLSLATNPKNDS 117
Qy 121 FMACAPLWSRACSSVSFSSGICARVDASFOQGSLLAPTAQRCPTYMDVTVLQSGNSIYP 180
Db 118 FLACSPWSHECSSYYTQMCSSRVNSFRFSKTVAPALQRCQYMDIVVLQSGNSIYP 177
Qy 181 WSEVQTELRRLVGLKFLIDPEQIQVGLVQGESVHEWMSLGDFTKEVVRVAAKNLSRREG 240
Db 178 WVEVQHEFLINLLKFFIGPQIQVGVVQYGEDVVEHFLNDYRSKDVVEASHIEQGG 237
Qy 241 RETKTAQIMVACTEGSPQSHGGRPEARLLVVVTGESHGDBELPAALKACAGRVTRY 300
Db 238 TETRTAFGIEFARSEAFQK--GGRKGAKVMIVITDGESHSDSPLEKVIQQSRDNTVRY 295
Qy 301 GIAVLGHVLRQRDPSSFLREIRTIASDPDEREFENVTDAAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYINRRGINPETFLNEIKYIASDDPDHFFNVTDAAALKDIVDALGDRIFSLGEGT 355
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYADMGSGVLMLEGHRLPPRMALBDEF 420
Db 356 N-KNETSFGLMSQTFSSHVEDVLLGAVGAYDMNGAVLKETSACKVPLRESYLKEP 414
Qy 421 PPALQNHAAVLYGSVSMILLRGGRRLLPSGAPRHRHGKVIATQKKDGVAVVAQSLQGE 480
Db 415 PEELKNHGAYLGVTVTSSVSRQGRVVVAGAPRPNHTGKVLFTMNNRSLTIHQAMRGQ 474
Qy 481 QIGSYFGSELCPDTRDGTDLVLLVAAPMLPGPQNKETGRVVVYLVGQOSSLTLOQTLLQ 540
Db 475 QIGSYFGSEITSDVIDDGDGVTDLVLLVGPAMYFN--EGRERKVVVYEL--RQNRFFVYNTLTK 532
Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDHQALYLYHGQTQGVRRPAPQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVVVVGAPLEDNHAGAIYIFHGRFSILKTPQRI 592
Qy 600 AASMPHALSYFGRSDVGRDLDDGLDVAVAGQGAAILSSRPYVHLTPSLVETVQAI 659
Db 593 TASELATGLQVFGSIHGQLDLNEDGLIDLAVGALGNVILWSPVQVQINASHLFFBPSKI 652
Qy 660 SVVQDCCRERGOEAVCLTAALCQVTSRTGRWDHQFYMTASLDTSWAGARAAPFGSG 719
Db 653 NIFHRDCKRGRDATCLAAFLCPTPIFLAPHFQTTVGIIRYNATMDERRRTPRAHLDEGG 712
Qy 720 QRLSPRLRLSVGNVTCQLHFHVLDTSDYLRPVALTVTFALDNTTKPGPVNLGSGPTSI 779
Db 713 DRFTNRAVLLSSGQELCERINFHVLDTADYVKPVTFSVYSLEDDPDH-GPMLDDGWPTTL 771
Qy 780 QKLVFPFKDQPDNECVTDLVLQVNMDOI-----RGRK-----APFVV 817

Db 772 RVSVFWNGCDEHCYVDLVLARSDDLPTAMEYQORVLRKPAQDCSAYTILSPDTTVFII 831
 Qy 818 RGRKRVLTLENNKKNAYNTSLTIFRNHLASLTQORSPKIVCEAAPS--AHAR 875
 Db 832 ESTQRVAVATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIECVNEBERRLOKQ 891
 Qy 876 LCSVGHPVFTQAKVTFLLEFEFSCSLLSQVFGKLTASSDSLRNCTLOENTAQTSAYI 935
 Db 892 VCNVSYFFRAKAVARLQDFEFSKIFLHHLELELAAGSDSNERDSTKEDNVAUFRHL 951
 Qy 936 QYBPHLLFSSESTLHYEVHPYGTLP--VGPGEFKTTLRVQNLGCVVSGLIISALLPA 993
 Db 952 KYEADVLFTRSSLSHVEKPNSSLYRVDGIGPPFCIFRIQNLGLPIHGMWKTIPI 1011
 Qy 994 VAHGNYFLSLQVITN--NASC-IVQNLTPEPPGVPHPELOHNLNLSNTOCQVRC 1050
 Db 1012 ATRSGNRLKLRDLTDEANTSCNIMGNSTYRPTVE--EDLRRAPQNLHNSDVVINC 1070
 Qy 1051 HLGQLAGTEVSGLLRVHNEPRRAKFKSLTVVSTFELGTBEGSVLQLTEASRSES 1110
 Db 1071 NI-RLVNPQENFHLGNLWLRSLKALKYKSMKIMVNAALQORPHSPFIFREEDPSQIV 1129
 Qy 1111 LEVVQTRPILISLWILGSLVGLGILLALLVFCILWKLGFPAHKKIPBEKREKLE 1166
 Db 1130 FEISKQEDWQPIWIVGSLGILLALLVALLVWKLGLGFRSAR---RRREPCLD 1181
 RESULT 14
 AAU14467
 ID AAU14467 standard; protein; 1188 AA.
 AC AAU14467;
 XX
 XX 24-OCT-2001 (first entry)
 DE Human novel protein #338.
 XX
 XX Human; novel protein; Antianemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiaesthetic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX
 XX Homo sapiens.
 OS
 XX
 XX W0200155437-A2.
 PN
 XX
 PD 02-AUG-2001.
 XX
 XX 25-JAN-2001; 2001WO-US002623.
 PF
 XX
 XX 25-JAN-2000; 2000US-00491404.
 PR
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX
 XX WPI; 2001-451939/48.
 DR
 XX
 XX N-PSDB; AAS22772.
 DR
 XX
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage.
 XX
 XX
 XX Example 4; Page 828-831; 894pp; English.
 PS
 XX
 XX The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides

are used to identify compounds which bind to the polypeptides.
 CC polynucleotides of the invention are used as probes and primers, for
 CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise
 CC antibodies/elicite an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC represents a protein of the invention
 XX
 XX Sequence 1188 AA;
 SQ
 Query Match 39.9%; Score 2438.5; DB 4; Length 1188;
 Best Local Similarity 43.2%; Pred. No. 1e-207;
 Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;
 1 MELPFTVTHLPFLVFTLGLCSFNLDEHPRLPFGPPEAFEGYSVLQHVGGQGRWMLVGA 60
 1 MDLPRLGLVAVAWALSLMPGFTDTFNMDTRKPRVPGSRTAFPGYTVQOHDISGNKWLVVGA 60
 61 PWDGPGSDRGDYYRCPVPGCAHNAKCAKHLGDLQYLGNSHHPAVNMHLGSLLETDDGG 120
 61 PLETNGYQKTDGVYKCPV---IHGNCNKLNLGRVTLNSVSRKDNMLRGLSLATNPKNDS 117
 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTYMDVWVLDGNSIYP 180
 118 FLACSLPWSHECGSSYYTTCWCSRVNSFRFSKTVAPALQRCQTYMDIVLDGNSIYP 177
 181 WSEVQTFRLRLVKLFIDPEIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSRREG 240
 178 WVEVQHFLINILKFKYIGPGQIQGVVQYGEDVVHFEHLNDYRSVKDWEAAASHIEQRGG 237
 241 RETKTAQAINVACTEGFSQSHGGRPEARLLVVVTVTDCGSHDGEELPAALKACEAGRTY 300
 238 TETRTAFGIEFARSEAFQK--GGRKGAKKYMIWITDGHSDSPLEKVIQOOSRDNTRY 295
 301 GIAVLGHLRRQRPSPFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
 296 AVAVLGYNNRGINPEIFLNEIKYIASDDPKHFFNVTDAAALKDIVDALGDRIFSLEGT 355
 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGAYDWGCVLWLEGHRLFPFRMALEDEF 420
 356 N-KNETSFGLMSQIGFSTHRLKDGILFGMVGAYDWGCVLWLEGHRLFPFRMALEDEF 414
 421 PPALQNHAAVLYGYSVSMILRGRRLLFLSCAPRHRGKVIATOLKKGAVRVAOSLQGE 480
 415 PEELKNHGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSLSUTHQAMRQ 474
 481 QIGSYFSGELCPDTRDGTDLVLLVAAFLPQNKETGRVYVYLVGQGSLLTQGLTQ 540
 475 QIGSYFSGEITSVDIDGDDGTVLLVAGAPMYFN--EGRERKVVYVEL--RQNFVYNGTLK 532
 541 PEPP--QDAREFGAMGALPDNLQDGFADVAVGAPLEDCHQAGLYLYHGTQSGVRPHPAQRI 599
 533 DSHSYQNAARFGSSIASVRDLNQDSYNDVVVVGAPLEDNHAGAIYIFHGRGSIKTPQRI 592
 600 AAASMPHALSYFGESVDGRDLDDGDDLVAVAGQAAILLSRPPIVHLTPSLEVTQAI 659
 593 TASELATGLQYFGCSIHGOLDLNEGLDLAVGALGNVAILWSRPVQVQINASLHFEPSKI 652
 660 SVVQDCRRRGQAVCLTAALCFQVTSRTSGRWDHQFYMRTASDLDEWTAGARAAFGSG 719
 653 NIFHRCKRSGRDNATCLAAFLCFTPIFLAHFQTTTGVIRYNATWDERRYTPRAHDEGG 712

Db 772 RVSVFPWNGCNEDEHCVFDLVLDDARSDLPTAMEYCORVLRKPAQDCSAYTLSTFTTTFII 831
Qy 818 RGRKRVLVSTTLNRKENAYNTSLIIFSRNLHLASLTPORESPIKVECAAPS--AHAR 875
Db 832 ESTRQVAVEATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSGIECVNEERRLOKQ 891
Qy 876 LCSVGHVFQTAGKVTFLLEPEFSCSLLSQVFGKLTASSDSLERNGLQENTAOQSAYI 935
Db 892 VCNVSYPPFRKAKAVAFRLDFEFKSI FLHHLEIELAAGSDSNERDSTKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSESTLHRYEVHPYGTLP--VGPGEFKTTLRVQNLGCYVVSGLIISALLPA 993
Db 952 KYEVDVLFTRSSLSHYEVKPNSSLERYDVGIGPPFCIFRIQNLGLFPIHGMMKTIPI 1011
Qy 994 VAHGNVFLSLSQVITN--NASC-IVQNLTEPPGPPVHPPELOHTNRNGSNTOCQVVR 1050
Db 1012 ATRSGNRLKLRLDFTDEANTSCNIWGNSTYRPTVE-EDLRRAPQLNHSNDSVWSINC 1070
Qy 1051 HLGOLAKGTEVSVGLRLVHNEFFERRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESL 1110
Db 1071 NI-RLVNPQENFHLGNLRLSKALKYKSMKIMVNAALQORQFHPFIFREEDPSRQIV 1129
Qy 1111 LEVVQTRPILISLWILIGSVLGGLLALLVFCCLKLGFPAHKKIPBEEKREEKLE 1166
Db 1130 FEISKQEDWQVPIWIIVGSTLGGLLALLVLAALWKLGFFSAR----RRREPGLD 1181

Search completed: April 6, 2005, 12:22:27
Job time : 154.588 secs

The [illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:17:21 ; Search time 145.761 Seconds
(without alignments)
2658.060 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106

Sequence: 1 MELPFVTHLPLVFLVLTGLC.....GFFAHKKIPBEERKEKLEQ 1167

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 1418010

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6092	99.8	1167	US-10-741-601-531	Sequence 531, App
2	6092	99.8	1177	US-10-741-601-532	Sequence 532, App
3	2439.5	40.0	1188	US-10-291-265-338	Sequence 338, App
4	2438.5	39.9	1188	US-10-291-265-810	Sequence 810, App
5	2429	39.8	1189	US-09-984-130-35	Sequence 35, Appl
6	2429	39.8	1189	US-09-984-130-35	Sequence 35, Appl
7	2429	39.8	1189	US-10-262-839-43	Sequence 4, Appl
8	2254	36.9	1034	US-09-984-130-43	Sequence 43, Appl
9	2254	36.9	1034	US-09-836-353A-43	Sequence 43, Appl
10	2244.5	36.4	1120	US-10-262-839-6	Sequence 6, Appl
11	1894.5	31.0	1180	US-09-918-715-307	Sequence 307, App
12	1893	31.0	1179	US-09-918-715-250	Sequence 250, App
13	1887	30.9	1151	US-09-984-130-103	Sequence 103, App

14	1887	30.9	1151	10	US-09-836-353A-103	Sequence 103, App
15	1880	30.8	1177	17	US-10-741-600-1161	Sequence 1161, App
16	1756	28.8	1181	14	US-10-741-600-1161	Sequence 2, Appl
17	1756	28.8	1181	15	US-10-295-027-1286	Sequence 1286, App
18	1756	28.8	1181	15	US-10-211-462-187	Sequence 187, App
19	1740.5	28.5	1148	16	US-10-872-198-147	Sequence 147, App
20	1727.5	28.3	707	9	US-09-764-870-313	Sequence 313, App
21	1727.5	28.3	707	14	US-10-125-540-313	Sequence 313, App
22	1720	28.2	1147	15	US-10-336-603A-42	Sequence 42, Appl
23	1131.5	18.5	1161	9	US-09-350-259-2	Sequence 2, Appl
24	1131.5	18.5	1161	10	US-09-891-943-2	Sequence 99, Appl
25	1126	18.4	1161	9	US-09-350-259-99	Sequence 99, Appl
26	1126	18.4	1161	10	US-09-891-943-99	Sequence 55, Appl
27	1124.5	18.4	1161	9	US-09-350-259-55	Sequence 55, Appl
28	1124.5	18.4	1161	10	US-09-891-943-55	Sequence 53, Appl
29	1109	18.2	1161	9	US-09-350-259-53	Sequence 53, Appl
30	1109	18.2	1161	10	US-09-891-943-53	Sequence 37, Appl
31	1108.5	18.2	1151	9	US-09-350-259-37	Sequence 37, Appl
32	1108.5	18.2	1151	10	US-09-891-943-37	Sequence 1088, App
33	1106	18.1	1170	17	US-10-741-600-1088	Sequence 2, Appl
34	1105	18.1	1170	9	US-09-945-265-2	Sequence 1, Appl
35	1105	18.1	1170	15	US-10-261-164-1	Sequence 624, App
36	1104.5	18.1	688	10	US-09-866-050A-624	Sequence 295, App
37	1103.5	18.1	1223	16	US-10-408-765A-295	Sequence 1086, App
38	1103.5	18.1	1223	17	US-10-741-600-1086	Sequence 46, Appl
39	1103	18.1	1155	9	US-09-350-259-46	Sequence 46, Appl
40	1103	18.1	1155	10	US-09-891-943-46	Sequence 130, App
41	1100.5	18.0	1145	16	US-10-872-198-130	Sequence 3, Appl
42	1085.5	17.8	1153	9	US-09-350-259-3	Sequence 1, Appl
43	1085.5	17.8	1153	10	US-09-902-481A-1	Sequence 3, Appl
44	1085.5	17.8	1153	10	US-09-891-943-3	Sequence 30, Appl
45	1085.5	17.8	1153	14	US-10-144-259-30	

ALIGNMENTS

RESULT 1
US-10-741-601-531
; Sequence 531, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01500
; CURRENT APPLICATION NUMBER: US/10/741.601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 531
; LENGTH: 1167
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-741-601-531

Query Match 99.8%; Score 6092; DB 16; Length 1167;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1164; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MELPFVTHLPLVFLVLTGLCSPFNLDHHPRLFGPPEAEFGYSVLQHVGGQRMVYGA	60
Db	1	MELPFVTHLPLVFLVLTGLCSPFNLDHHPRLFGPPEAEFGYSVLQHVGGQRMVYGA	60
Qy	61	PWDGSGDRRGDVYRCVPGGAHNAAPCAKGLDYLQGNSSHPAVNMHLGMSLLETGDDGG	120
Db	61	PWDGSGDRRGDVYRCVPGGAHNAAPCAKGLDYLQGNSSHPAVNMHLGMSLLETGDDGG	120
Qy	121	FMACAPLWSRACGSSVSSGICARVDASFQPGSLAPTAQRCPTVMDVVIIVLDGNSNIYP	180
Db	121	FMACAPLWSRACGSSVSSGICARVDASFQPGSLAPTAQRCPTVMDVVIIVLDGNSNIYP	180
Qy	181	WSEVQTFELRLVGLKFLDPEQIQVLQVQYGSFVHWSLGLDFRTKEEVVRAAKNLSREG	240

```
Db 181 WSEVQTLRLRLVGLFDPEIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRREG 240
Qy 241 RETKTAQIMVACTEGFSQSHGGPEARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 241 RETKTAQIMVACTEGFSQSHGGPEARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLMLEGGHRLFPFRMALEDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLMLEGGHRLFPFRMALEDEF 420
Qy 421 PPALONHAAVLYGYSVSMLLRGGRFLSGAPRFRHGKVIATQFKKDGAVRVAQSLQGE 480
Db 421 PPALONHAAVLYGYSVSMLLRGGRFLSGAPRFRHGKVIATQFKKDGAVRVAQSLQGE 480
Qy 481 QIGSYFGSELCPDLTDGDTTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOSLLTLQGTQ 540
Db 481 QIGSYFGSELCPDLTDGDTTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOSLLTLQGTQ 540
Qy 541 PEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAQRIA 600
Db 541 PEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAQRIA 600
Qy 601 AASMPHALSVFGRSVGDRLLDGGDLVDVAVAGAAIILSSRPIVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSVFGRSVGDRLLDGGDLVDVAVAGAAIILSSRPIVHLTPSLEVTPOAIS 660
Qy 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Db 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVINEGSPTSIQ 780
Db 721 RLSPRRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVINEGSPTSIQ 780
Qy 781 KLVPFSKDCGPDNECVTDLVQNMDIRGSRKAPFVVRGGRRKVLVSTTLNKENAYNT 840
Db 781 KLVPFSKDCGPDNECVTDLVQNMDIRGSRKAPFVVRGGRRKVLVSTTLNKENAYNT 840
Qy 841 SLSLIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPFQTKAKVTFLLEPFSC 900
Db 841 SLSLIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPFQTKAKVTFLLEPFSC 900
Qy 901 SLLSQVFGKLTASSDSLERNGLTOBNTAOTSAYIQVEPHLLFSSES9TLHRYEYVHPTGL 960
Db 901 SLLSQVFGKLTASSDSLERNGLTOBNTAOTSAYIQVEPHLLFSSES9TLHRYEYVHPTGL 960
Qy 961 PVGPGPEFTKTLRVNQLGCVVSGLIISALLPAVAGHGNFSLISOVITNNASCIVQNL 1020
Db 961 PVGPGPEFTKTLRVNQLGCVVSGLIISALLPAVAGHGNFSLISOVITNNASCIVQNL 1020
Qy 1021 EPPGPPVHPELOKTNRLNGSNTQCVVRCHLGQALAKGTEVSGLLRLVHNEPFRRAKFK 1080
Db 1021 EPPGPPVHPELOKTNRLNGSNTQCVVRCHLGQALAKGTEVSGLLRLVHNEPFRRAKFK 1080
Qy 1081 SLTVVSTFELGTEBGSVLQTEASRWSESLLEVVTQTPILISLWILGSLGALLLALL 1140
Db 1081 SLTVVSTFELGTEBGSVLQTEASRWSESLLEVVTQTPILISLWILGSLGALLLALL 1140
Qy 1141 VFCLWKLGFPAHKKIPREEKKEKLEQ 1167
Db 1141 VFCLWKLGFPAHKKIPREEKKEKLEQ 1167
```

RESULT 2

US-10-741-601-532

; Sequence 532, Application US/10741601

; Publication No. US20040166519A1

; GENERAL INFORMATION:

```
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-10-741-601-532
```

```
Query Match 99.8%; Score 6092; DB 16; Length 1177;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1164; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MELPFFVTHLFLPLVFLTGLCSPFNLDHHPRLFPFGPPEAEFGYSVLQHVGGQRMWLVGA 60
Db 1 MELPFFVTHLFLPLVFLTGLCSPFNLDHHPRLFPFGPPEAEFGYSVLQHVGGQRMWLVGA 60
Qy 61 PWDGPGSDRRGDVYRCFVGGAHNA PCAKGHLGDYQLGNSHHPAVNMHLGMSLLETGDDG 120
Db 61 PWDGPGSDRRGDVYRCFVGGAHNA PCAKGHLGDYQLGNSHHPAVNMHLGMSLLETGDDG 120
Qy 121 FMACAPLWSRACSSVFSSGICARVDASFPQGS LAPTAQRCPTYMDVVIVL DGSNSIYP 180
Db 121 FMACAPLWSRACSSVFSSGICARVDASFPQGS LAPTAQRCPTYMDVVIVL DGSNSIYP 180
Qy 181 WSEVQTLRLRLVGLFDPEIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRREG 240
Db 181 WSEVQTLRLRLVGLFDPEIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRREG 240
Qy 241 RETKTAQIMVACTEGFSQSHGGPEARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 241 RETKTAQIMVACTEGFSQSHGGPEARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLMLEGGHRLFPFRMALEDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLMLEGGHRLFPFRMALEDEF 420
Qy 421 PPALONHAAVLYGYSVSMLLRGGRFLSGAPRFRHGKVIATQFKKDGAVRVAQSLQGE 480
Db 421 PPALONHAAVLYGYSVSMLLRGGRFLSGAPRFRHGKVIATQFKKDGAVRVAQSLQGE 480
Qy 481 QIGSYFGSELCPDLTDGDTTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOSLLTLQGTQ 540
Db 481 QIGSYFGSELCPDLTDGDTTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOSLLTLQGTQ 540
Qy 541 PEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAQRIA 600
Db 541 PEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAQRIA 600
Qy 601 AASMPHALSVFGRSVGDRLLDGGDLVDVAVAGAAIILSSRPIVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSVFGRSVGDRLLDGGDLVDVAVAGAAIILSSRPIVHLTPSLEVTPOAIS 660
Qy 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Db 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVINEGSPTSIQ 780
Db 721 RLSPRRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVINEGSPTSIQ 780
Qy 781 KLVPFSKDCGPDNECVTDLVQNMDIRGSRKAPFVVRGGRRKVLVSTTLNKENAYNT 840
Db 781 KLVPFSKDCGPDNECVTDLVQNMDIRGSRKAPFVVRGGRRKVLVSTTLNKENAYNT 840
```

```
Qy 841 SLILISRNHLASLTPORESPIKVECAAPSAHARLCSVGHVPVQTKAKVTFLEPFESC 900
Db 841 SLILISRNHLASLTPORESPIKVECAAPSAHARLCSVGHVPVQTKAKVTFLEPFESC 900
Qy 901 SSLLSOVFGKLTASSDSLENGTLQENTAOISAVIQVEPHLLPSESTLHRYEHPYGTLL 960
Db 901 SSLLSOVFGKLTASSDSLENGTLQENTAOISAVIQVEPHLLPSESTLHRYEHPYGTLL 960
Qy 961 PVGPGPEFKTTLRVQNLGCVYVSGLIISALLPAVAHGNFYLSLSQVITNASCIVQNL 1020
Db 961 PVGPGPEFKTTLRVQNLGCVYVSGLIISALLPAVAHGNFYLSLSQVITNASCIVQNL 1020
Qy 1021 EPPGPPVHPELOHTNRLNGSNTQCVVRCHLQOLAKGTESVGLLRLVHNEFPRAKFK 1080
Db 1021 EPPGPPVHPELOHTNRLNGSNTQCVVRCHLQOLAKGTESVGLLRLVHNEFPRAKFK 1080
Qy 1081 SLTVVSTFELGTREGSVLQTEASRWSESLLEVVTQTRPILISLWILGSLVGLGLLLALL 1140
Db 1081 SLTVVSTFELGTREGSVLQTEASRWSESLLEVVTQTRPILISLWILGSLVGLGLLLALL 1140
Qy 1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167
Db 1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167
```

RESULT 3

```
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338
```

```
Query Match 40.08; Score 2439.5; DB 15; Length 1188;
Best Local Similarity 43.28; Pred. No. 1.1e-206;
Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;
```

```
Qy 1 MELPFTVTHLPLVFLTGLCSPFNLDHHPRLPFPPEAEFGYSVLQHGQGRWMLVGA 60
Db 1 MDLPRGLVVAWALSFWGFTDTFMDTRKPRVIFGSRAPFGYTVQGHDSGNKWLVGA 60
Qy 61 PWGPGSDRRGDVYRCVPGGAHNAAPCAKGLGQYQLGNSHPAVNMHLGMSLLETGDGG 120
Db 61 PLETNGYQKTDVYKCPV---IHGNCCTKLNLGRVTLNVSERKDNMRGLSLATNPKNDS 117
Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTMYDMVIVLDSNSIYP 180
Db 118 FLACPLWSHCEGSSYTTGMCRSWNFSRFTKVPALQRCQTYMDIVIVLDSNSIYP 177
Qy 181 WSEVOTFLRLVLKGLFTDPEQIQGLVQYGPSVHWSLGDFTKKEVWRAAKNLSREG 240
Db 178 WVEQVHFLINILKFIYIGPQIQGVVQYQGVGDDVHVEFHLNDYRSVKDVEAASHIEQGG 237
```

RESULT 4

```
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
```

```
Qy 241 RETTAAQIMVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKACAGRVTRY 300
Db 238 TETRTAFIEFARSEAPQK--GGRKGAKKVMIWITGESHSDSPLEKVIQOOSRDNTRY 295
Qy 301 GIAVLGHLRRQRDPSPSFLREIRTIASDPDEREFNFVNTDEAALTDIYDALGDRIFGLEGS 360
Db 296 AVAVLYGYNRGINPETFLNEIKYIASDPDDKHFFNVTDAAALKDIYDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMEQIGFSTHRLKDGILFGMWGAYDGGSVLMLEGGHRLPPRMALEDEF 420
Db 356 N-KNETSFGLMEQIGFSSHVVEDGVLGAVGAYDMNGAVLTKETSAGKVIPLRESYLKEF 414
Qy 421 PPALQNHAAVLYGYSVSMMLRGRRLLFSGAPRHRGKVIAFOLKKGAVRVAQSLQGE 480
Db 415 PSELKHGALVGYTVTSVSSROGRVAVAGAPRNFHTGKVLFTMHNRRSLTHQAWRGQ 474
Qy 481 QIGSYFGESELCPLDTRDGTDLVLAAMPFLGPQNKETGRVYVYLQGQSLTLQGTTLQ 540
Db 475 QIGSYFGESEITSVDIDGDTVLLVGAPWYFN-EGREKGVYVEL-RQNRFYVNGTLK 532
Qy 541 PEPQ-QDARPGFAMGALPDNLQDGFADVAGAPLEDHGHQALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNAREGSSIASVRDLNQDSYNDVVVGCAPLEDNHAGAIYIFHGFSGSILKTPQRI 592
Qy 600 AAASMPHALSYFGRSVDRLDGDGLVDVAVGAQGAAILLSRPIVHLTPSLEVTPOAI 659
Db 593 TASELATGQYFGCSIHGQDLNEDGLIDLAVGALGNVILMRPVVQINASLHFEFSKI 652
Qy 660 SVVQDCRRRGQBAVCLTAALCFQVTSRTPGRDHGFYMRFTASLDDEWTAGARAAFQSG 719
Db 653 NIFHRCKRSGRDATCLAAFLCTPIPLAFHQTTTVGIRYNATMDERRYTPRAHLDGG 712
Qy 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTPTFALDNTTKPGVNLGSPSTSI 779
Db 713 DRETNRALLSSGOELCERINFVLDTADYVVKPVTFSEYSLEDPDH-GPMLDDGWPTTL 771
Qy 780 QKLVPFKSKDCPDNECVTDLAVQNMDI-----RGRK-----APVV 817
Db 772 RVSPVFWNGCNEDEHCVDFDLVDARSDLPTAMEYCORVLRKPAQDCSAYTSLSDTTVFII 831
Qy 818 RGGRRKVLVSTTLNKRKENAYNTSLSIIFSRNLHLASLTPORESPIKVECAAPS--AHAR 875
Db 832 ESTQRVAVATEATLENGENAYSVLNISQSANLQFASLIQKEDSDGSEICVNEERRLOKQ 891
Qy 876 LCSVGHVPVQTKAKVTFLEPFESCSLLSOVFGKLTASSDSLENTQTLQENTAOISAYI 935
Db 892 VCNVSVPPFRKAKAVAPRLDFEFSKIFLHLELELAAGSDSNRSDSTKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSESTLHRYEHPYGTLP--VGPGEFKTTLRVQNLGCVYVSGLIISALLPA 993
Db 952 KYEADVLFTRSSLSLSHYEVPNSLERYDGGPPFCIFRIQNLGLPFIHGMKMITIPI 1011
Qy 994 VAHGNYFLSLSQVITN--NASC-IVONLTPEPPVHPELOHTNRLNGSNTQCVVRC 1050
Db 1012 ATSGHRLKLRPLDTEANTSCNIWGNSTYEPTPVE-EDLRAPQLNHSNDVVSINC 1070
Qy 1051 HLGQLAKGTESVGLLRLVHNEFPRAKFKSLTVVSTFELGTREGSVLQTEASRWSESL 1110
Db 1071 NI-RLVNPQINFHLLGNLWRLSLKALKYKSMKIMVNAALQRPSPPIFREEDPSQIV 1129
Qy 1111 LEVVQTRPILISLWILGSLVGLLLALLVFLWKLGLGFAHKKIPEEKREKLE 1166
Db 1130 FEISKQBDWQVPIIIVGSTLGGLLALLVFLWKLGLGFAHKKIPEEKREKLE 1181
```

```
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 810
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-291-265-810

Query Match      39.8%; Score 2438.5; DB 15; Length 1188;
Best Local Similarity 43.2%; Pred. No. 1.4e-206;
Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;

Qy 1 MELPFVTHLPLVLVLTGLCSFNLDEHHPLRPPGPEAEFGYSLVQHVGQGRWMLVGA 60
Db 1 MDLPRGLVVAWALSMPGFTDTFNMTRKPRVTPGSRPTAFPGYTVQOHDISGNKWL VVGA 60
Qy 61 PWGPGSDRGDGVYRCVPGGAHNAAPCAKAGHLGDLQYLGNSHPAVNMHLGSLLETDDGG 120
Db 61 PLETNGYQKTGDYVKCPV---IHGNCCTKLNLGRVTLNSVSRKDNMRGLSLATNP KDNS 117
Qy 121 FMACAPLWSRACSSVFSSGICARVDASFOQGS LAPTAQRCPTYMDVWIVL DGSNIYP 180
Db 118 FLACSPLWSHCEGSSYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVL DGSNIYP 177
Qy 181 WSEVQTFRLRLVKGFLIDPBIQVGLVQYGESPVHWSLGDPRTKKEVVRVAAKNLSRREG 240
Db 178 WVEVQFELNLIKFKYIGPGQIQGVVQYGEDVVHPEFLNDYRSVDVVEAASHIEQRGG 237
Qy 241 RETTAQAIIVACTEGFSQSHGGRPEARLLVVVTDGESHGDELPAALKACENGRVTRY 300
Db 238 TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOOSERDNTRY 295
Qy 301 GIAVLGHLRRORPSSFLREIRTIASDPERRFPFNTDEAALTIDVDALGDRIFGLEGS 360
Db 296 AVAVLGYNNRGINPETFLEIKYIASDDPKHFPNTDEAALKDIDVDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMSQIGFSFTHRLKDGILFGVMGVADMGSVLWLEGHRLFPFPRMALEDEF 420
Db 356 N-KNETSFGLMSQTFSSHVVEDGVLGAVGAYDMGAVLKETSAKGVIPLRSEYLKEF 414
Qy 421 PPALONHAAYLGYSVSSMLRGGRLFLSGAPRFRHKGVIAPQLKDGAVRVAQSLQGE 480
Db 415 PEELNKHGALGYTVTVSVSSRQGRVYVAGAPRHNHTGKVILFTMHNRSLLTHQAMRGG 474
Qy 481 QIGSYFSGELCPDTRDGTDLVLLVAAPFLGPQNKETGRVYVVLVGGQSLLTLOQTLO 540
Db 475 QIGSYFSGEITSVIDGSDGTVDLVVGAPMYFN--EGRGRGVYVEL--QNRFPYNGTLK 532
Qy 541 PEPP-QDARFGFANGALPDNLQDGFADVAAGAPLEDGHQGALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVWVGAPLEDNHAGAIYIFHGFSGSILKTPKQRI 592
Qy 600 AAASMPHALSYFGSVNDRDLDDDLVDVAVAGAGAAILLSRPIVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGGSIHQDLNDEGLIDLAVGALGNAVLMSRPVVQINASLHFEPSKI 652
Qy 660 SVVORDCRRRQEQEAVCLTAALCFQVTSKTPGRWDHQFYMRFTASLDSEWTAGARAAPDGG 719
Db 653 NIFHRDCKSRGDATCLAAFLCTFPIFLAHPQTTTIGIRYNATMDERRYTTPRAHLEGG 712
Qy 720 QRLSPRLRLRISGVNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVNLSEGSTSI 779
Db 719 QRLSPRLRLRISGVNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVNLSEGSTSI 779

; SEQUENCE 35, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-984-130-35

Query Match      39.8%; Score 2429; DB 10; Length 1189;
Best Local Similarity 43.1%; Pred. No. 9.8e-206;
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

Qy 1 MELPFVTHLPLVLVLTGLCSFNLDEHHPLRPPGPEAEFGYSLVQHVGQGRWMLVGA 60
Db 1 MDLPRGLVVAWALSMPGFTDTFNMTRKPRVTPGSRPTAFPGYTVQOHDISGNKWL VVGA 60
Qy 61 PWGPGSDRGDGVYRCVPGGAHNAAPCAKAGHLGDLQYLGNSHPAVNMHLGSLLETDDGG 120
Db 61 PLETNGYQKTGDYVKCPV---IHGNCCTKLNLGRVTLNSVSRKDNMRGLSLATNP KDNS 117
Qy 121 FMACAPLWSRACSSVFSSGICARVDASFOQGS LAPTAQRCPTYMDVWIVL DGSNIYP 180
Db 118 FLACSPLWSHCEGSSYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVL DGSNIYP 177
Qy 181 WSEVQTFRLRLVKGFLIDPBIQVGLVQYGESPVHWSLGDPRTKKEVVRVAAKNLSRREG 240
Db 178 WVEVQFELNLIKFKYIGPGQIQGVVQYGEDVVHPEFLNDYRSVDVVEAASHIEQRGG 237
Qy 241 RETTAQAIIVACTEGFSQSHGGRPEARLLVVVTDGESHGDELPAALKACENGRVTRY 300
Db 238 TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOOSERDNTRY 295
Qy 301 GIAVLGHLRRORPSSFLREIRTIASDPERRFPFNTDEAALTIDVDALGDRIFGLEGS 360
Db 296 AVAVLGYNNRGINPETFLEIKYIASDDPKHFPNTDEAALKDIDVDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMSQIGFSFTHRLKDGILFGVMGVADMGSVLWLEGHRLFPFPRMALEDEF 420
Db 356 N-KNETSFGLMSQTFSSHVVEDGVLGAVGAYDMGAVLKETSAKGVIPLRSEYLKEF 414
Qy 421 PPALONHAAYLGYSVSSMLRGGRLFLSGAPRFRHKGVIAPQLKDGAVRVAQSLQGE 480
Db 415 PEELNKHGALGYTVTVSVSSRQGRVYVAGAPRHNHTGKVILFTMHNRSLLTHQAMRGG 474
Qy 481 QIGSYFSGELCPDTRDGTDLVLLVAAPFLGPQNKETGRVYVVLVGGQSLLTLOQTLO 540
Db 475 QIGSYFSGEITSVIDGSDGTVDLVVGAPMYFN--EGRGRGVYVEL--QNRFPYNGTLK 532
Qy 541 PEPP-QDARFGFANGALPDNLQDGFADVAAGAPLEDGHQGALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVWVGAPLEDNHAGAIYIFHGFSGSILKTPKQRI 592
Qy 600 AAASMPHALSYFGSVNDRDLDDDLVDVAVAGAGAAILLSRPIVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGGSIHQDLNDEGLIDLAVGALGNAVLMSRPVVQINASLHFEPSKI 652
Qy 660 SVVORDCRRRQEQEAVCLTAALCFQVTSKTPGRWDHQFYMRFTASLDSEWTAGARAAPDGG 719
Db 653 NIFHRDCKSRGDATCLAAFLCTFPIFLAHPQTTTIGIRYNATMDERRYTTPRAHLEGG 712
Qy 720 QRLSPRLRLRISGVNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVNLSEGSTSI 779
Db 719 QRLSPRLRLRISGVNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVNLSEGSTSI 779

; SEQUENCE 35, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-984-130-35

Query Match      39.8%; Score 2429; DB 10; Length 1189;
Best Local Similarity 43.1%; Pred. No. 9.8e-206;
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

Qy 1 MELPFVTHLPLVLVLTGLCSFNLDEHHPLRPPGPEAEFGYSLVQHVGQGRWMLVGA 60
Db 1 MDLPRGLVVAWALSMPGFTDTFNMTRKPRVTPGSRPTAFPGYTVQOHDISGNKWL VVGA 60
Qy 61 PWGPGSDRGDGVYRCVPGGAHNAAPCAKAGHLGDLQYLGNSHPAVNMHLGSLLETDDGG 120
Db 61 PLETNGYQKTGDYVKCPV---IHGNCCTKLNLGRVTLNSVSRKDNMRGLSLATNP KDNS 117
Qy 121 FMACAPLWSRACSSVFSSGICARVDASFOQGS LAPTAQRCPTYMDVWIVL DGSNIYP 180
Db 118 FLACSPLWSHCEGSSYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVL DGSNIYP 177
Qy 181 WSEVQTFRLRLVKGFLIDPBIQVGLVQYGESPVHWSLGDPRTKKEVVRVAAKNLSRREG 240
Db 178 WVEVQFELNLIKFKYIGPGQIQGVVQYGEDVVHPEFLNDYRSVDVVEAASHIEQRGG 237
Qy 241 RETTAQAIIVACTEGFSQSHGGRPEARLLVVVTDGESHGDELPAALKACENGRVTRY 300
Db 238 TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOOSERDNTRY 295
Qy 301 GIAVLGHLRRORPSSFLREIRTIASDPERRFPFNTDEAALTIDVDALGDRIFGLEGS 360
Db 296 AVAVLGYNNRGINPETFLEIKYIASDDPKHFPNTDEAALKDIDVDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMSQIGFSFTHRLKDGILFGVMGVADMGSVLWLEGHRLFPFPRMALEDEF 420
Db 356 N-KNETSFGLMSQTFSSHVVEDGVLGAVGAYDMGAVLKETSAKGVIPLRSEYLKEF 414
Qy 421 PPALONHAAYLGYSVSSMLRGGRLFLSGAPRFRHKGVIAPQLKDGAVRVAQSLQGE 480
Db 415 PEELNKHGALGYTVTVSVSSRQGRVYVAGAPRHNHTGKVILFTMHNRSLLTHQAMRGG 474
Qy 481 QIGSYFSGELCPDTRDGTDLVLLVAAPFLGPQNKETGRVYVVLVGGQSLLTLOQTLO 540
Db 475 QIGSYFSGEITSVIDGSDGTVDLVVGAPMYFN--EGRGRGVYVEL--QNRFPYNGTLK 532
Qy 541 PEPP-QDARFGFANGALPDNLQDGFADVAAGAPLEDGHQGALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVWVGAPLEDNHAGAIYIFHGFSGSILKTPKQRI 592
Qy 600 AAASMPHALSYFGSVNDRDLDDDLVDVAVAGAGAAILLSRPIVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGGSIHQDLNDEGLIDLAVGALGNAVLMSRPVVQINASLHFEPSKI 652
Qy 660 SVVORDCRRRQEQEAVCLTAALCFQVTSKTPGRWDHQFYMRFTASLDSEWTAGARAAPDGG 719
Db 653 NIFHRDCKSRGDATCLAAFLCTFPIFLAHPQTTTIGIRYNATMDERRYTTPRAHLEGG 712
Qy 720 QRLSPRLRLRISGVNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVNLSEGSTSI 779
Db 719 QRLSPRLRLRISGVNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTKPGPVNLSEGSTSI 779
```

Db 118 FLACPLWSHECCSSYYTTCGSRVNSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Qy 181 WSEVQTFRLRLVKGKLFIDPEQIQVGLVQYGESPVHWSLGDPRFKTEVVAARKNLSRREG 240
Db 178 WBEVQHFILNLIKFKYIGPQIQGVVQYGEDVHVEPHLNDYRSKDVVEAASHIEQGG 237
Qy 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 238 TETRTAFGIEFARSEAPQK--GGRKGAKYMIIVITGESHSDSPLEKVIQOOSRDNVTRY 295
Qy 301 GIAVLGHLRRORDPSPFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLYGNRRGINPETFLNEIKYIASDPDKHFFNVTDAAALKDVIDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDNGGSLVLEGGHRLFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQIGFSSHWEDGVLGAGVAYDMGAVLKETSAGKVIPLRESYLKEF 414
Qy 421 PPALQNHAAVLYGSVSMMLRGRRLLFLSGAPFRHGRKVIATFOLKKGAVRVAQSLQGE 480
Db 415 PEBLKNHGAYLGYVTVTSVSSROGRVYVAGAPRNFHTGKVLFTMHNRRSLTIHQAWRGQ 474
Qy 481 QIGSYFGSELCPLDTRDGTDLVLLVAAPMFLGPQNKETGRVYVLYVGOQSLTLTQGLTQ 540
Db 475 QIGSYFGSEITSDVIGDGTVDVLLVGAPMYFN--EGRERKVVYVEL--RQNRFYVNGTLK 532
Qy 541 PEPP--QDARFGAMGALPDINQDGFADVAVGAPLEDGHQAGLYLHYGTQSGVRPHPAQRI 599
Db 533 DSHSYQNAFGSSIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIFHGFGRSILKTPKQRI 592
Qy 600 AAASMPHALSYFGRSVUDGRDLDDGLDVLVAVGAQGAAILLSRPPIVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSPVQVQINASHLHFESKI 652
Qy 660 SVVQDRCRRRQGAVALCLTAALCFQVTSRTGRWDHQFYMRFTASLDSEWTAGARAAFDGSG 719
Db 653 NIHRDCKRSGRDATCLAAFLCTPIPLAFHFQTTTGVIRYNATMDERRITPRAHLDGEG 712
Qy 720 QRLSPRLRLSVGNVTCQELHFHVLDTSDYLRPVALTFTFALDNTTKPGPVNAGSPSTI 779
Db 713 DRFTNRAVLSSGOELCERINFHVLDTADYVVKPVTFSVEYSLEDPDH--GPMLDGDMPTTL 771
Qy 780 QKLVPSKDCPDNECVTDVIVQNMDI-----RGRK-----APFVV 817
Db 772 RVSPVPMWNGCNEDEHCVPLDLVLDARSOLDPTAMEYQCRVLRKPAQDCSAYTLSDFTTVFII 831
Qy 818 RGRKRLVSTTLLENKENAYNTSLIIFSRNLHLASLTQRESPIKVECAAPS--AHAR 875
Db 832 ESTRQRAVAVATLENGENAYSTVLNISQSANLQFASLIQKEDSDGSEICWNERRLQKQ 891
Qy 876 LCSVGHVPVFGAKVTFLLBEPFSCSLLSQVFKLTASSDSLSERNGTLOQNTAQTSAIY 935
Db 892 VCNVSYFPFRAKAVAPRLDPEFSKSIFLHLEIETELAAGSDSNERDSTKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSSESTLHRYEVHPYGTLP--VGPGEFPTTLRVONLGCYVSVGLIISALLPA 993
Db 952 KYEADVLFTSSLSHYEVKLSLERYDYGIPFPFCIFRIQNLGLFPIHGMKMITIPI 1011
Qy 994 VAHGNYFLSLQVITN--NASC-IYONLTERRPPVPHPELOHTNRLNGSNTOCQVVR 1049
Db 1012 ATRSNRLKLRPLDDEVANTSCNIWGNSTERYPTVE--EDLRAPQLNHSNDVVSIN 1070
Qy 1050 CHLGQLAGKTEVSVGLLRLVHNEFFRRAKFSLTVSTFELGTBEGSVLQLEASRWSES 1109
Db 1071 CNI--RLVPNQEINPHLLGNLWRLSLKALKYKSMKIMVNAALQRFHSPFIFREEDPSRQI 1129
Qy 1110 LLEVQTRPILISLWILGVLGGLLALLVFCWLKGLFPFAHKKIPBEERKREBLE 1166
Db 1130 VFEISKQEDQVPTIIVGSTGLGGLLALLVLAALWKLGFPRFAR----RRRPEGLD 1182

; Sequence 35, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PP489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-836-353A-35

Query Match 39.8%; Score 2429; DB 10; Length 1189;
Best Local Similarity 43.1%; Pred. No. 9.8e-206;
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

Qy 1 MELPFTVTHLPFLPVLTGLCSPFNLDHHPRLPPGPEAEFGYSVLQHVGGQGRWMLVGA 60
Db 1 MDLPRLGLVAVAWALSLLWPFGTDTFNMOTRKPRVTPGSRTAFFGYTVQQHDSGNKWLAVGA 60
Qy 61 PWDGPGSDRRGDVYRCPVGGAHNAPCAKGHGLGYQLGNSHPAVNHLGMSLLETDCDGG 120
Db 61 PLETNGYQKTDVYKPCPV---IHGNCCKLNLGRVTLNSVSRKDNMRLGLSLATNPKNDS 117
Qy 121 FMACAPLWSRACSSVFSFGICARVDASFOQGLSAPTAQRCPTYMDVWVLDGNSIYP 180
Db 118 FLACPLWSHECCSSYYTTCGSRVNSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Qy 181 WSEVQTFRLRLVKGKLFIDPEQIQVGLVQYGESPVHWSLGDPRFKTEVVAARKNLSRREG 240
Db 178 WBEVQHFILNLIKFKYIGPQIQGVVQYGEDVHVEPHLNDYRSKDVVEAASHIEQGG 237
Qy 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 238 TETRTAFGIEFARSEAPQK--GGRKGAKYMIIVITGESHSDSPLEKVIQOOSRDNVTRY 295
Qy 301 GIAVLGHLRRORDPSPFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLYGNRRGINPETFLNEIKYIASDPDKHFFNVTDAAALKDVIDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDNGGSLVLEGGHRLFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQIGFSSHWEDGVLGAGVAYDMGAVLKETSAGKVIPLRESYLKEF 414
Qy 421 PPALQNHAAVLYGSVSMMLRGRRLLFLSGAPFRHGRKVIATFOLKKGAVRVAQSLQGE 480
Db 415 PEBLKNHGAYLGYVTVTSVSSROGRVYVAGAPRNFHTGKVLFTMHNRRSLTIHQAWRGQ 474
Qy 481 QIGSYFGSELCPLDTRDGTDLVLLVAAPMFLGPQNKETGRVYVLYVGOQSLTLTQGLTQ 540
Db 475 QIGSYFGSEITSDVIGDGTVDVLLVGAPMYFN--EGRERKVVYVEL--RQNRFYVNGTLK 532
Qy 541 PEPP--QDARFGAMGALPDINQDGFADVAVGAPLEDGHQAGLYLHYGTQSGVRPHPAQRI 599
Db 533 DSHSYQNAFGSSIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIFHGFGRSILKTPKQRI 592
Qy 600 AAASMPHALSYFGRSVUDGRDLDDGLDVLVAVGAQGAAILLSRPPIVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSPVQVQINASHLHFESKI 652
Qy 660 SVVQDRCRRRQGAVALCLTAALCFQVTSRTGRWDHQFYMRFTASLDSEWTAGARAAFDGSG 719
Db 653 NIHRDCKRSGRDATCLAAFLCTPIPLAFHFQTTTGVIRYNATMDERRITPRAHLDGEG 712

[illegible]

RESULT 8

```

RES001.8
US-09-984-130-43
; Sequence 43, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-43

```

RESULT 9
US-09-836-353A-43
; Sequence 43, Application US/09836353A
; Publication No. US20030129685A1


```

; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 6
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-262-839-6

Query Match          36.4%; Score 2224.5; DB 15; Length 1120;
Best Local Similarity 40.6%; Pred. No. 1.3e-187;
Matches 485; Conservative 199; Mismatches 397; Indels 115; Gaps 17;

QY 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFFGPPPEAFGYSVLQHVGGGQRMVLVGA 60
Db 1 MDPLRGLVWAALSLWPGFTDTFNMDTRKPRVPGSRTAFPGYTVQQHDSGNKWL VVGA 60
QY 61 PWDGSGDRGDYVRCVGGAAHNAACPAGKHGLDYGQGNSSHPAVNMHLGMSLLETDDGG 120
Db 61 PLETNGYQKTDGVYKCEV-----IHGNC TKJNLG----- 89
QY 121 FMACAPLWSRACGSSVFSGGICARVDASFPQGS LAPTAQRCPTYMDVIVLDGNSIYP 180
Db 90 -----CQTYMDIVIVLDGNSIYP 108
QY 181 WSEVQTELRRLVGLKFLDPDQIQVLQYQGESPVHWSLGDGFRTKBEEVRAAKNLSRREG 240
Db 109 WVEVQHFLINILKKFYIGPGQIQGVGVQYGEDVVHEFLNDYRSVKDVEAASHIEQGG 168
QY 241 RETKTAQAIWVACTEGFSQSHGRPEARLLVVVTDGESHDEGELPAALKACEAGRVTY 300
Db 169 TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGESHSDPLEKVIQOOSERDNTY 226
QY 301 GIAVLGHLRRORDPSSFLREIRTIASDPDRPFNVTDEAALTDIVDALGDRIFGLEGS 360
Db 227 AVAVLGYNRRGINPETFLNEIKYIASDPDDKGFNVTDEALXDIVDALGDRIFLSLEGT 286
QY 361 HAENESSFGLEMSQIGFSTHRLKDGIILFGVMGAYDWGGSVLWLEGGHRLFPFRMALEDEF 420
Db 287 N-KNETSFGLEMSQTGFSHVVEDGVLLGAVGAYDWNGAVLKETSGAKVPIRESYLKEF 345
QY 421 PPALQNHAAVLYGYSVSNMLRGGRRRLFLSAGPRPHRGKVITAPOLKXGDGAVRVAQSLOGE 480
Db 346 PEELEKNHGAYLGYTVTSVWSRQGRVTVAGAPRENHTGKVLFTMHNRSLSITHQAMRQG 405
QY 481 QIGSYFSGELCPJLTDTRDGGTVDVLLVAAPFLGPKQNETGRVYTVLVGQQSLLTQCTLIQ 540
Db 406 QIGSYFSGEISITSDVIDGGVTVDLVLLGAPMYFN--EGRERGKVYYVEL--RQNRVFNVTGLK 463
QY 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDGQGAALYHGTQSGVRPHPAQRI 599
Db 464 DSHSYQNRFGSSIASVVDLNDQSYNDVVVGAPLEDNHAGAIYFHGFRGSILKTPQRI 523
QY 600 AAASMPHALSYFGRSVDRDLDDGDLVDVAVGAAQGAAILSSRPPIVHLTFSLVTPQAI 659
Db 524 TASELATGLQYFGCSIHQGLDNLNEDGLDILAVGALGNNAVILWSRPVQVINASLHFEPSKI 583
QY 660 SVVORDCRRRGOEAVCLTAALCFQVTSRTTCGRWDHQFVMRPTASLDSEWTAGARAAFDSG 719
Db 584 NI FHRDCGRGRDATCLAAFLCFTPI FLAPHFQTTVTGIRYNAITMDERRYTPRAHLDEGG 643
QY 720 QRLSPRRLRLSVGNVTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLNEGSPSIT 779
Db 644 DRNTNRAVLSSGQELCERINFHVLDTADYVKPFTFSVEYSLEDPDH--GPMLLDGGWTTLL 702
QY 780 QKLVPPFSKDCGPDNCEVCTDVLVQNMDDI-----RGSRK-----APFVV 817
;

```


QY 774 GSPTSIOKLVFPKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGGRKVLVSTTLNLR 833
DB 793 SLFNSVHEYIPFAKDCGCKEKCISDLHLV-----ATTEKOLLIVRSQNDKFNVSITVKNT 848
QY 834 KENAYNTSLGIIIFSRNLHLASLTQRESPIKVECAAPSAAHARLCSVGHVPFOTGAKVTEL 893
DB 849 KDSAYNTRTVHVSFNLVFSGI-----EAIQKDS-ESNHNITCKVGYFLRRGEMVTFK 902
QY 894 LEFEFSSLLSQVFGKLTASSDSLRNGTLQENTQTSAYIOYEPHLLSSSESTLHRYE 953
DB 903 ILFOFNTSYLMENVTIYLSATSDSEPPETLSDNVNINISIPVKYEVGLQFVSSASEVHIS 962
QY 954 VHPYGTLP-----VGPGEPEKTLRVQNLGCVVSGLIISALLPAVAHGN---YFLSL 1004
DB 963 IAAETVPEVINSTEDIGNEINIFYLRKSGSPMPPELKLISFPNMTSGPYVLYPTGL 1022
QY 1005 SQVITNASCIVQNLTEP-----PGPPVHPPEELOHTNRLNGSTQCVVRCHLGQAKG 1058
DB 1023 SS--SENANCRPHIFEDPFINSKMTTSTDHLKRGITLDCNTCKEATTICNLTLS-SDI 1079
QY 1059 TEVSUGLRLVHNEFFRRKPKSLTVVSTFELGTEGSLQLTEASRWSLELLEVQOT-R 1117
DB 1080 SQNVSL--ILWKPTFKSYFSSNLNLTIRGELSENAS-LVLSSSNQKRELAIQISKDGL 1136
QY 1118 PILISLWILGSLVGLGILLALLVFCVWKLGFPAHKKIPEEKREE 1163
DB 1137 PGRVPLVILLSAFAGLLMLLILALWKIGFF---KRPLKXKMEK 1179

RESULT 13
US-09-984-130-103
; Sequence 103, Application US/09984130
; Publication No. US2003005231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: P48992
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR FILING DATE: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR FILING DATE: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR FILING DATE: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR FILING DATE: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR FILING DATE: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-103

Query Match 30.98; Score 1887; DB 10; Length 1151;
Best Local Similarity 36.38; Pred. No. 1.4e-157;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;

QY 23 FNLDHHPRLFPFGPEAEFGYSLQHVGGGQRMWLVGAPMDPGSGDRRGDYVRCVGAH 82
DB 1 FNVDKNSMTFSGVEDMFGYVQYENESGKWLIGSLVLPVQPKNTGDIYKCPVGRGE 60
QY 83 NACPACGHLGDYQVGNSSHPAV-----NMHLGMSLLETDDGGGPMACAPLWSRACGSSVF 137
DB 61 SLPCVKLDLP-----VNTSIPNVTVEKENMTFG-STLVTPNPGGFLACGPLYAYRCGHLHY 115
QY 138 SSGICARVDASFOQGSILAPTAQRCPTYMDVIVLDCGNSIYPHSEVQTFLLRLVGLFI 197
DB 116 TTGICSDVSPTFQVNVNIAIP-VQECSTQLDIVIVLDCGNSIYPWDSVTAFLNDLLKRMDI 174

RESULT 14
US-09-836-353A-103
; Sequence 103, Application US/09836353A

QY 198 DPQIQVGLVQYGESPVHWSLGDFTKKEVVRRAAKNLSRRREGRTTKTAQAIMVACTEGP 257
DB 175 GPKQTOVGIQYQGENVTHEFNANKYSSTEEVLVAAKKIVQGGGQNTALGTDTPARKEAF 234
QY 258 SQSHGGRPEARLLVVVYDGSDDGELPAALKACEAGRVTRGIAVLGHVLRQRDPSS 317
DB 235 TEARGARGVGVVWIVDGEDSHNHLKVKVQCEDENIQRFSLILGSYNRNLSTEK 294
QY 318 FLREIRTIASDPDERFFNFVNTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGF 377
DB 295 FVEEIKSIASEPTEKHFNFVSDDELALVTIVKTGLERIFALEATADQSAASFEMEMSQTF 354
QY 378 STHRLKDGILFGWAGYDWCWSVLWLEGHRLPPRMALEDEPPALQNHAYLYGYSVSS 437
DB 355 SAHYSQDWMVLGAVGADWNGTVMQKASQIIIPRNTTFNVESTKQNEPLASYLYGTVNS 414
QY 438 MLLRGGRRLFLSGAPRHRGKVIAFOLKKOGAVRAQSLQGEQIGSYFSGELCPLDTR 497
DB 415 ATASSGDVLIAGQPRYHHTGVIIYRM-EDGNIKILQILSGEIQIGSYFSGILTTTDDK 473
QY 498 DGTDTLLVLAAPMLGPGQNKETGRVYVYLVGQOSSLTLOGTLQ----- 541
DB 474 DSNTDILLVGAAPMVGTEKEEQKVVYAL-NQTRFVQMSLEPIKOTCCSSROHNSCTT 532
QY 542 ---EPPODARFGFAMGALPDINODGFADVAVGAPLEDHOGALYLYHGTQSGVPHPAQR 598
DB 533 ENKNEPCGARFGTAIAVAKDLNLDGFNDIVIGAPLEDHGGAVVYIHGSGKTIKRYAQR 592
QY 599 IAAASPHALSYFGRSVDGRLDGDLDLVAVAGCAAGAAILLSSRPVHLTPSLVTPQA 658
DB 593 IPSGGDKTLKFFGQS IHGEMDNLGDLTDTVITGGLGAALFWSRDVAVVAVKVTWNFENK 652
QY 659 ISVVQRDCRRRGOEAVCLTAALCFQVTSRTPGRMDHOFYMRFTASLDDEWAGARAAFDGS 718
DB 653 VNIQKNCHMEGETVCINATVCFEVLKSKEDTIYADLQYRVTLDSLQISRSFSGT 712
QY 719 GQRLSPRRLSLVGNVTCEQLHFVLDTSYLRPALVTFTALDNTTKP--GPVLNEGSP 776
DB 713 QERKQVR--NITVRKSECTKHSFVMLDKHDFQDSVRITLDF--NLTDPNGPVLDDSLP 767
QY 777 TSJQKLVPSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGGRKVLVSTTLNENKEN 836
DB 768 NSVHEYIIPFAKDCGCKEKCISDLHLV-----ATTEKOLLIVRSQNDKFNVSITVKNTKDS 823
QY 837 AYNTSIIIFSRNLHLASLTQRESPIKVECAAPSAAHARLCSVGHVPFOTGAKVTFLLF 896
DB 824 AYNTRTIVHVSFNLVFSGI-----EAIQKDS-ESNHNITCKVGYFLRRGEMVTFKILF 877
QY 897 EFSCSSLLSQVFGKLTASSDSLRNGTLQENTQTSAYIOYEPHLLSSSESTLHRYEVHP 956
DB 878 QFNTSYLMENVTIYLSATSDSEPPETLSDNVNINISIPVKYEVGLQFVSSASEVHISAA 937
QY 957 YGTLP-----VGPGEPEKTLRVQNLGCVVSGLIISALLPAVAHGN---YFLSLSQV 1007
DB 938 NETVPEVINSTEDIGNEINIFYLRKSGSPMPPELKLISFPNMTSGPYVLYPTGLSS- 996
QY 1008 ITNASCIVQNLTEP-----PGPPVHPPEELOHTNRLNGSTQCVVRCHLGQAKGTEV 1061
DB 997 -SENANCRPHIFEDPFINSKMTTSTDHLKRGITLDCNTCKEATTICNLTLS-SDISQV 1054
QY 1062 SVGLLRLVHNEFFRRKPKSLTVVSTFELGTEGSLQLTEASRWSLELLEVQOT-RPIL 1120
DB 1055 NVSL--ILWKPTFKSYFSSNLNLTIRGELSENAS-LVLSSSNQKRELAIQISKDGLPGR 1111
QY 1121 ISLWILGSLVGLGILLALLVFCVWKLGFPAHKKIPEEKREE 1163
DB 1112 VPLWILLSAFAGLLMLLILALWKIGFF---KRPLKXKMEK 1151

```
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR FILING DATE: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-103

Query Match      30.8%; Score 1887; DB 10; Length 1151;
Best Local Similarity 36.3%; Pred. No. 1.4e-157;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;

Qy 23 FNLDLHPRLPFGPPPEAEFGYSLVQHVGGQRMWLVGAPWDPGSGDRRGDVYRCPVCGAH 82
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1 FNVDDKNSMTFSGPVEDMFGYTVQYENESGKWLIGSLVGPQPKNTGDIYKCPVGRGE 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 83 NAPCAKGLHDYQLGNSHPAV-----NMHLGMSLLETGDDGGMACAPWRSACGSSVF 137
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 61 SLPCVKLDLP-----VNTSIPNVTEVKENMTFG-STLVTPNPGGFLACGPLYAVRCGHLHY 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 138 SSGICARVDASFOQGSGLAPTAQRCPYMDVIVLDGNSNITYPMSEYQTLRLVGLFI 197
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 116 TTGTICSDVSTFQVQNSIAP-VQECSTQLDIVLDGNSNITYPMDSVTAFLNDLLKEMDI 174
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 198 DPEIQVGLVQYQVGSPPVHESGLDGRFRTKEEVRAAKNLSRREGRETKTAQAIMVACTEGF 257
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 175 GPKQTQVGIQYQVGENVTHEFNLNKYSSTEEVLVAAKIVQRGGRQWTALGTDTARKEAF 234
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 258 SQSHGGRPEARLLVVTDGESHDPGEELPAALKAACEAGRVTRYGIAGVILGHYLRQRDPSS 317
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 235 TEARGARGVKVMVITDGEHDHNRLLKKVIOQCEDENIQRFSAIILGSYNRGNLSTEK 294
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 318 FLREIRTIADPDPRFFPNVTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGF 377
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 295 FVEIKSIASEPTKHFNFVSDLEALVTIVKTLGERIPALEATADQSAASFEMESQTGP 354
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 378 STHRLKDGILFGMYGAYDWGQSVLWLGHRFLFPPRMALEDEFPALQNHAAVILGYSVSS 437
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 355 SAHYSQDWMLGAVGADWNGTVVMQKASQIIIPNTTFNVESTKKNEPLASYLGYTVNS 414
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 438 MLRGGRRLLFUSGAPRFRHKGKVIJAFOLKXGAVRVAQSLQGEQIGSYFSGELCPDTR 497
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 415 ATASSGDVLVYAGQPRYNHTQVQIYIRM-EDGNKILQTLGSGEIQIGSYFSGILTTDIDK 473
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 498 DGTDDVLLVAAPMFLPQNKETGRVYVYLVCQSLTLTLOGLP----- 541
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 474 DSNTDILLVGAPMYMGTEKEEGKVVYAL-NQTRFYQMSLEPIKQTCSSSRQHNSCTT 532
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 542 ---BPPQDARFGFANGALPOLNQDFADVAVGAPLEDGCHQALYLYHGTQSGVPHPAQR 598
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 533 ENKNEPCARGRTAIAAVKOLNLDGFNDIVIGAPLEDHGHGAVIYHSGKTIKREVAQR 592
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 599 IAAASMPHALSYFGRSVDRLDLDGDLVDVAVAGQAAILLSRRPIVHLTFPSLEVTPQA 658
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 593 IPSGGDGKTLKFFQSGIHGEMDLNGDGLTDVTIGLGLGAALFWSRDVAVVAVVKNFEPNK 652
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 659 ISVVQRCCRGRQAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAPDGS 718
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 653 VNIOKNCHMEGKETVCINATVCPEVKLSKEDTIYADLUQYRVTLDSLRQISRSFSGT 712
```

RESULT 15

```
US-10-741-600-1161
; Sequence 1161, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1161
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-600-1161
```

```
Query Match      30.8%; Score 1880; DB 17; Length 1177;
Best Local Similarity 36.3%; Pred. No. 6e-157;
Matches 430; Conservative 225; Mismatches 455; Indels 76; Gaps 23;
```

```
Qy 20 CSPFNLDLHPRLPFGPPPEAEFGYSLVQHVGGQRMWLVGAPWDPGSGDRRGDVYRCPVG 79
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 26 CVSNFVDDKNSMTFSGPVEDMFGYTVQYENESGKWLIGSLVGPQPKNTGDIYKCPVG 85
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 80 GAINAPCAKGLHDYQLGNSHPAV-----NMHLGMSLLETGDDGGMACAPWRSACGS 134
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 86 RGSGLPCVKLDLP-----VNTSIPNVTEVKENMTFG-STLVTPNPGGFLACGPLYAVRCGH 140
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 135 SVFSSGICARVDASFOQGSGLAPTAQRCPYMDVIVLDGNSNITYPMSEYQTLRLVGLK 194
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 141 LHYTTGTICSDVSTFQVQNSIAP-VQECSTQLDIVLDGNSNITYPMDSVTAFLNDLLER 199
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 195 LFTDPEIQVGLVQYQVGSPPVHESGLDGRFRTKEEVRAAKNLSRREGRETKTAQAIMVACT 254
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 200 MDTPGKQTQVGIQYQVGENVTHEFNLNKYSSTEEVLVAAKIVQRGGRQWTALGTIDTARK 259
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 255 EGFSSQSHGGRPEARLLVVTDGESHDPGEELPAALKAACEAGRVTRYGIAGVILGHYLRQRD 314
```

Db 260 EAFTEARGARGVKKVWVIITDGHSHNHLKVKIOCEDENIQRFSAIALGYNRNL 319
Qy 315 PSSFLREIRTIASDPDRFPFNTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQ 374
Db 320 TEKFVEIKSIASEPTKEHFNVSDELALVTIVKTGLERIFALEATADQSAASFEMMSQ 379
Qy 375 IGFSHRLKDGILFGMVGVADWGSVUWLGHHRLFPFRMALEDEPPPALONHAAYLGYS 434
Db 380 TGFSAHYSQDWMVGAVGDMGTVMQKASQIIIPRNTTFNVESTKQNEPLASVILGYT 439
Qy 435 VSSMLLRGRRFLSGAPRFRHGRKVIATFOLKKGAVRVAQSGOIGSYFGSELCPD 494
Db 440 VNSATASSGVLYIAGQPRNHTGQVYIYRM-EDGNIKILQILSGEOIGSYFSGILTTD 498
Qy 495 TDRDGTDDVLLVAAPMPLGPNKQKTRGVVYVVGQQLTLQGTLOP----- 541
Db 499 IDKSNNTDILLVGPATMGTKEBQGVVYVAL-NQTRFEYQMSLEPIKQTCSSRQHNS 557
Qy 542 -----EPPQDARFGFAMGALPDINQDGFADVAVGAPLEDHOGALYLYHGTSGVVRP 595
Db 558 CTTENKNEPCGARFGTAIAVAKOLNLDGFNDIVIGAPLEDHGGAVYIYHSGKTKRKEY 617
Qy 596 AQRIAAASMPHALSYFGRSVDGLDLDGDDLVAVGCAQGAAILSSRPVHLTPSLEVT 655
Db 618 AQRIPSGGDKTKLPFGQS IHGEMDLNGDGLTDVTICGLGGAALFWSRDAVAVKVTWNE 677
Qy 656 POAISVVQRDCRRRGQAVCLTALCFQVTSRTIPGRWDHOFYRFTASLDEBTAGAPAT 677
Db 678 PNKVNIOKKCHMEGKETVCINATVCFDVKLSKEDTIYEADLQYRTV 677
Qy 716 DSGQRLSPRRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTVTFALDNTTKP--GPV 773
Db 738 SGTQERKQVR--NITVRKSECTKHSFYML--HDPQDSVRITLDF--NLDPENGPVLD 790
Qy 774 GSPTSIOQLVPFSKCGPDNECVTDVLQVNMDIRGSKAPFVVRGGRKVLSTTLNR 833
Db 791 SLPSNVHEYIPFAKDCGCKEKCSDSLHV-----ATTEKOLLIVRSQNDKFNVS 846
Qy 834 KENAYNTSLIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVFTGAKV 893
Db 847 KDSAYNTRTIVHSPNLVPSGI----BAIQKDS--ESNHNITCKVGYPLRGEWTFK 900
Qy 894 LEFEFCSLLSQVFGKLTASSDSLERNGTQNTAQTSAIYQYEPHLLFSSBSTLHRYE 953
Db 901 ILFQNTSYLMENVTIYLSATSDEEPPETLSDNVVNISIPVKYEVGLQFVSSASEYHIS 960
Qy 954 VHPYGTLP-----VGPPEKTLTVONLGCYVWSGLIISALLPAVAGGN---YFLSL 1004
Db 961 IAAETVPEVINSTEDIGNEINIFYLIRKSGFPMPKLSISFPNMTSNGYPVLYPTGL 1020
Qy 1005 SQVITNNASCIVQNLTEP-----PGPPVHPEELQHTNRLNGSNTQCQVVRCHLQ 1058
Db 1021 SS--SENANCRPHIFEDFPFINSCKMTTSDHLKRGITLDCNTCKFATITCNLTS-SDI 1077
Qy 1059 TEVSVGLRLRVHNEFFRFRAXFKSLTVVSTFELGTEGSLVQLTEASRWSLSLEVQOT-R 1117
Db 1078 SQNVSL--ILWKPFTFKSVFSSNLRTIRGELRSENAS-LVLSSNQKRELAIQISKDGL 1134
Qy 1118 PILISLWILGVLGHLILLALLVFLWKLGFPAHKKIPEEKREE 1163
Db 1135 PGRVPLWILLSAFAGULLMLLALLWKGIF---KRPLKKQMEK 1177

Search completed: April 6, 2005, 12:53:00
Job time : 151.761 secs

This Page Intend (empty)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 49.8054 Seconds
(without alignments)
1749.117 Million cell updates/sec

Title: US-09-647-544-2
Perfect score: 6106
Sequence: 1 MELPFVTHLFLPLVFLTGLC.....GFFAHKKIPREKREKLEQ 1167

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pap:*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pap:*
3: /cgn2_6/ptodata/1/iaa/6A-COMB.pap:*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pap:*
5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pap:*
6: /cgn2_6/ptodata/1/iaa/backfiles.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2429.5	39.8	1217	4	US-09-949-016-7892
2	1894.5	31.0	1180	4	US-09-000-004A-2
3	1784	29.2	1183	4	US-09-532-310B-5
4	1756	28.8	1181	4	US-09-000-004A-4
5	1756	28.8	1181	4	US-09-949-016-6189
6	1756	28.8	1195	4	US-09-949-016-10747
7	1728	28.3	1183	4	US-09-532-310B-6
8	1131.5	18.5	1161	1	US-08-173-497-2
9	1131.5	18.5	1161	1	US-08-286-889-2
10	1131.5	18.5	1161	1	US-08-485-618-2
11	1131.5	18.5	1161	1	US-08-362-652-2
12	1131.5	18.5	1161	2	US-08-605-672-2
13	1131.5	18.5	1161	2	US-08-482-293A-2
14	1131.5	18.5	1161	2	US-08-943-363-2
15	1131.5	18.5	1161	3	US-08-193-043-2
16	1131.5	18.5	1161	4	US-09-688-307A-2
17	1131.5	18.5	1161	4	US-09-350-259-2
18	1126	18.4	1161	1	US-08-485-618-99
19	1126	18.4	1161	2	US-08-605-672-99
20	1126	18.4	1161	2	US-08-482-293A-99
21	1126	18.4	1161	2	US-08-943-363-99
22	1126	18.4	1161	3	US-09-193-043-99
23	1126	18.4	1161	4	US-09-688-307A-99
24	1126	18.4	1161	4	US-09-350-259-99
25	1124.5	18.4	1161	3	US-09-193-043-55
26	1124.5	18.4	1161	4	US-09-688-307A-55
27	1124.5	18.4	1161	4	US-09-350-259-55

28	1118.5	18.3	1161	1	US-08-485-618-55	Sequence 55, Appl
29	1118.5	18.3	1161	1	US-08-362-652-55	Sequence 55, Appl
30	1118.5	18.3	1161	2	US-08-605-672-55	Sequence 55, Appl
31	1118.5	18.3	1161	2	US-08-482-293A-55	Sequence 55, Appl
32	1118.5	18.3	1161	2	US-08-943-363-55	Sequence 55, Appl
33	1109	18.2	1161	1	US-08-485-618-53	Sequence 53, Appl
34	1109	18.2	1161	1	US-08-362-652-53	Sequence 53, Appl
35	1109	18.2	1161	2	US-08-605-672-53	Sequence 53, Appl
36	1109	18.2	1161	2	US-08-482-293A-53	Sequence 53, Appl
37	1109	18.2	1161	2	US-08-943-363-53	Sequence 53, Appl
38	1109	18.2	1161	3	US-09-193-043-53	Sequence 53, Appl
39	1109	18.2	1161	4	US-09-688-307A-53	Sequence 53, Appl
40	1109	18.2	1161	4	US-09-350-259-53	Sequence 37, Appl
41	1108.5	18.2	1151	1	US-08-286-889-37	Sequence 37, Appl
42	1108.5	18.2	1151	1	US-08-485-618-37	Sequence 37, Appl
43	1108.5	18.2	1151	1	US-08-362-652-37	Sequence 37, Appl
44	1108.5	18.2	1151	2	US-08-605-672-37	Sequence 37, Appl
45	1108.5	18.2	1151	2	US-08-482-293A-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1

US-09-949-016-7892
; Sequence 7892, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7892
; LENGTH: 1217
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-7892

Query Match		39.8%	Score 2429.5	DB 4	Length 1217
Best Local Similarity		43.1%	Pred. No. 1.6e-225		
Matches 516		Conservative 213	Mismatches 422	Indels 45	Gaps 16
Qy	1	MELPFVTHLFLPLVFLTGLCSPFNLDHHPRLFPGPPEAEFGYSLVGHVGGQRMWVGA	60		
Db	30	MDLPRGLVVAWALSFWFGFTDTFNMTRKPRVIFGSRTPFGYTVQQHDISGNKWLWGA	89		
Qy	61	PWDGPGDRRGGVYRCPVGAHNAAPCAKHGLDYQLGNSHHPAVNMHGLMSLLETGDGG	120		
Db	90	PLETNGYKGTGDVYKCPV---IHGNCITKLNLGRVTLNSVSRKDNMRGLSLATNPXDNS	146		
Qy	121	FMACAPIWSRACGSSVFFSGICARVDASFQPGSLAPTAQRCPTMDVVIDGNSNIYP	180		
Db	147	FLACSPWSEHCSSYTTGMSRVNSFRFSKTPALQRCQTYMDIVIVLDGNSNIYP	206		
Qy	181	WSEVQTLRLRLVGLFIDPEQIQVGLVQYGSPPVHWSLGFRTKEEVVRAAKNLSRREG	240		
Db	207	WVEVQHFLINILKFLIPGQIQGVVQYQGVGVVHFNLDYRSVKDVVEAASHIEQRCG	266		
Qy	241	RETQTAAQIMVACTEGFSQSHGRPEARLLVVVVTGDSHDEELPALKACEAGRVTY	300		
Db	267	TEITRTAFGIEFARSEAFQK--GGRKGAKVMIVITDGESHDSPLDKVIOQSRDNVTY	324		
Qy	301	GIAVLGHYLRQRDPSSFLIRTIASDPDRFFNVNTDEAALTDIVDALGDRIFGLEGS	360		

```
Db 325 AVAVLYGNRRGINPETFLNEIKYIASDDPKHFNVTDEAALKDIVDALGDRIFSLEGT 384
Qy 361 HAENSEFGLGEMSGIQTSTHRLKDGILPGWGVADWGGSVLMLEGGHRLPPRMALEDEF 420
Db 385 N-KNETSFGLGEMSGTSSHVVDGVLGAVGADWNGAVKETSAGKVIPLRESYLKEF 443
Qy 421 PPALQNHAAVLGYSVSMILRGGRRLFLSGAPRHRGKVIAFOLKXDGAVRVAQSLQGE 480
Db 444 PEELKNHGAVLYTTSVSSRQGRVYVAGAPRHTGKILFTMHNNRSLTHQAMRGQ 503
Qy 481 QIGSYFGESECLPDLTDRTDGLVLLVAAPMLPGQNKETGRVYVYLVGQQSLTLQOTLQ 540
Db 504 QIGSYFGESEITSVBIDGDTVDLLVGAPMYFN-EGRERGKVVYEL-RQNRFPVYNGTLK 561
Qy 541 PEPP-QDARFGFANGALPDNLQGFADVAVGAPLEDHCHQALYLYHGTQGVRRPRAORI 599
Db 562 DSHSYQARFGSSIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIFHGFSGSILTKPKORI 621
Qy 600 AAASMPHALSYFGRSVDRGLDLDGDDLDVAVAGQAAILLSRPVHLTPSLEVTPOAI 659
Db 622 TASELATGLQYFGCSINGQLDLEDGLDLAVGALGNVILMSRPVVOINASHLFEPSKI 681
Qy 660 SVVORDCRRRQGEAVCLTAALCFQVTSRTFGWMDHOFMYMFTASLDEBWTAGARAADGSG 719
Db 682 NIFHRDCKRSGRDATCLAAFLCTPIFLAPHFQTTTGVIRYNATMDERRVTPRAHLDGG 741
Qy 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRVALVTFTALDNTTKPGVILNBSPTS 779
Db 742 DRFTNRAVLSSGQELCERINFHVLDTADYVKPFTSVBSYLEDPOH-GPMLDDGWPTTL 800
Qy 780 QKLVFPKSGDGPONECVTDLVQVNMDI-----RGRK-----APFVV 817
Db 801 RVSPFPGNCGNEHCVPDLVDARSDLPTAMEYQVRVLRKPAQDCSAYTSLSPDTTVFII 860
Qy 818 RGRKRVLVSTTLNRKENAYNTSLIIFGRNLHLASLTPORESPIKVECAAPS--AHAR 875
Db 861 ESTRQRAVATELENGENAYSTVNLISQSANLQFASLIQKSDSGSIECVNEERRLOKQ 920
Qy 876 LCSVGHVPFGTAKVFTLLFEFSCSLLSQVQFKLTASSDSLERNCTLQENTAQTSAYI 935
Db 921 VCNVSYFFRAKAVAPRLDFEFKSIFLHLELELAAGSDSNERDSTKEDNAPLRFHL 980
Qy 936 QYEPHLLFSSSESTLHRYEHPYGTLP--VGPGPEFKTTLRVQNLGCVVSGLIISALLPA 993
Db 981 KYEADVLPTRSSLSHVEVKNLSLERYDGGPPFCIPRIQNLGLFPIHGMWKITPI 1040
Qy 994 VAHGGNYFLSLSQVITN--NASC-IVQNLTPEPGPPVHPELOHTNRLNGSNTCCQVVR 1050
Db 1041 ATRSGNRLKLRLDPLTDDEANTSCNMGNSTEYRTPVE-BDLRRAPQLNHSNDVVSINC 1099
Qy 1051 HLQOLAGTQSVSGLLRLVNEPFRRAKFKSLTVVSTFELGTREGSVLQLTEASRWSEL 1110
Db 1100 NI-KLVNQINPHLLGNLWRLSLKALKYMKIMVNAALQROPHSPFFIFREBDSQV 1158
Qy 1111 LEVVQTRPILISLWILGSLVGLLILALLVFCILWKLGFPAHKKIPEEKREZKLE 1166
Db 1159 FEISKQEDWQVPIWIVGSTLGLLALLLVALLVWLKLGFRSAR-----RRREFGLD 1210
```

RESULT 2

```
US-09-000-004A-2
; Sequence 2, Application US/09000004A
; Patent No. 6780603
; GENERAL INFORMATION:
; APPLICANT: Talsibary, Photini-Effie
; APPLICANT: Charonis, Aristidis S.
; APPLICANT: Setty, Suman
; APPLICANT: Mauer, Michael
; TITLE OF INVENTION: ANALYSIS OF ALPHA INTEGRINS FOR THE DIAGNOSIS OF DIABETIC NEPHRO
; FILE REFERENCE: 600.314USWO
; CURRENT APPLICATION NUMBER: US/09/000.004A
; CURRENT FILING DATE: 2001-05-02
```

```
; PRIOR APPLICATION NUMBER: US 60/001.387
; PRIOR FILING DATE: 1995-07-21
; PRIOR APPLICATION NUMBER: US 60/001.861
; PRIOR FILING DATE: 1995-08-03
; PRIOR APPLICATION NUMBER: US 60/016.700
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: PCT/US96/12067
; PRIOR FILING DATE: 1996-07-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Rattus
; US-09-000-004A-2

Query Match 31.0%; Score 1894.5; DB 4; Length 1180;
Best Local Similarity 36.5%; Pred. No. 1.1e-173;
Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;

Qy 13 LVFLTGLCSPPNLDHHPRLFPQPPAEFGYSVLQHVGGQRMVLVGAWDGSGDGRGD 72
Db 19 LTVILGFCVSFNVVKNMSFSFGPVEDMFGYTVQYENEEGKWLIGSLVVGQPKARTGD 78
Qy 73 VYRCPVGGAHNAPCAKGHLDYQLGNSHPAV-----NMHLGMSLLETDGCGFMACAPL 127
Db 79 VYKCPVGRERAMPCKLDLP-----VNTSIPNVTIENKNTFG-STLVTPNGGFLACGPL 133
Qy 128 WSACSSVFSGICARVDASFOQSLAPTACRCPTYMDVVIVLDGNSIYPMWSVQTF 187
Db 134 YAYRCGLHYLTGICSDVSPTFOVNSFAP-VQECSTQLDIVIVLDGNSIYPMWSVIAF 192
Qy 188 LRRLVKGLFIDPQIOVGLVOYGESPVHWSLQDPTKESVVRPAAKNLSRREBRETAKQ 247
Db 193 LNDLLKRMIDIGPKQTQVIGVQYGVNTEFNLNKYSSTEELVVAANKIQRQGLQMTAL 252
Qy 248 AIMVACTEFSOSHGRPEAARLLVVVTDGESHGDELPAALKACBAGRVTRYGIADVGH 307
Db 253 GIDTARKEAFTEARGARGVKVWVIVTDGESHNVRLKQVQDCEDENIQRSIALGH 312
Qy 308 YLRQRDPSFARIEITIASDPDERFFNVVTDEAALTDIVDALGDRIFGLGSHARNES 367
Db 313 YNRGNLSKTEKFEIKSIASEPTKHPFNVSDELALVTIVKALGERIFALEATADQSAAS 372
Qy 368 FGLEMSQIGFSTHRLKDGILPGWGVADWGGSVLMLEGGHRLPPRMALEDEPPALQNH 427
Db 373 FEMESQTGFSAHYSQDVMYLGAVGAYDWNGTVMQKANOQMVIPHNTTFQTE--PAKONE 430
Qy 428 --AAYLGYSVSSMLLGRRLFLSGAPRFRHRGKVIAFOLKXDGAVRVAQSLQGEQIGSY 485
Db 431 PLASLYGTVNSATIPGD-VLYIAGQPRYNHTQOVVIYKM-EDGNINILQTLGGEQIGSY 488
Qy 486 FGSELCPDLTDRTDGLVLLVAAPMLPGQNKETGRVYVYLVGQQSLTLQOTLQ- 541
Db 489 FGSLVLTIDIDKDSYTDLLVGAPMYMTGKEBQGVVYVAV-NQTRFEYQMSLEPIROT 547
Qy 542 -----EPPQDARFGFANGALPDNLQGFADVAVGAPLEDHCHQALYLYHG 586
Db 548 CSSLKDNCKTENKNEPCGARFGTAIAAVKOLNVDFNDFVIGAPLEDHAGAVYIYG 607
Qy 587 TQSGVRPHPAQRIAAASMPHALSYFGRSVDRGLDLDGDDLDVAVAGQAAILLSRPV 646
Db 608 SGTITREAVAQRIPSGGDGKTLKFFQOSIHGEMDNGDGLTDVTIIGLGAALFWARDVA 667
Qy 647 HLTPSLEVTPQAISVVQDRCRRRQGEAVCLTAALCFQVTSRTFGWMDHOFMYMFTASLDE 706
Db 668 VVKVTWNFEPNKVINIQKNCRVKEGTVCINATMCFPHVKLSKEDSIYEADLQYRVTLDS 727
Qy 707 WTAGARAAFDGSGQRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRVALVTFTALDNTTK 766
Db 728 LRQISRSFSGTQERKIQIR-NITVRESECIHRHSFYMLDKHDFQDSVRVTLDFE---NLTD 782
Qy 767 P--GPVLNBSGPTSIQKLVFPKSGDGPONECVTDLVQVNMDIRGSRKAPFVVGRGRKV 824
```

Db 783 PENGVLDDALPNSVHEHIPPACOGKNCERCISDLTLNVST-----TEKSLLIIVKSQDKXF 838
Qy 825 LVSTTLNKRKENAYNTSLSIIFSRNLHLASTP-QRESPIKVECAAPSAHARLCSVGHVP 883
Db 839 NVSLTVKNKGDASYNTRTVQHSNPLIFSGIEEIQDS-----CESNQNTCRVGVYF 891
Qy 884 FQTGAKVTFLEPFSCSSLLSQVFGKLTASSDLENGTLQENTAOISAVIQEYPHLLP 943
Db 892 LRAGETVTFKIIFOFNTSHLSNAIIHLSATSDSEEPLESLENDNEVNAISIPVKYEVLQF 951
Qy 944 SSESTLHRYEHPYGTLP-----VGPGEPEKTLRVQNLCCYVVSGLIISALLP-AVAH 996
Db 952 YSSASEHHISVAANETIPEFNTSTEDIGNEINLVYTIKRGHPMPPELOLSISFNLITAD 1011
Qy 997 GGNVFLSLOVITNASCIVQNLTEPPG-----PPVHPEELOHTNRLNGSNTQCQVVR 1050
Db 1012 GYPVLYPIGWSSDNVNCPRSLDPFGINSKGKMTISKSEVLKRGTIQDCSSTCGVATI 1071
Qy 1051 HLGOLAKG-TEVSVGLRLVHNEFFRRAKFKSLTVSTFELGTGEGSVLQLTEASRWSES 1109
Db 1072 TCSLLPSDLQVNVSL--LLWKPTFIRAHFSSLLTLRGLKSENS--LTLSSNRKREL 1128
Qy 1110 LLEVQVQ-RPTLISLWILIGVLGGLLLALLVFLCWLKLGFAHKKIPPEEKREE 1163
Db 1129 AIQSKDGLQGVPLWILLSAFAGLLMLLLIALLWKIGFF---KRPLKKKMEK 1180

RESULT 3

US-09-532-310B-5
; Sequence 5, Application US/09532310B
; Patent No. 6596276
; GENERAL INFORMATION:
; APPLICANT: Senger, Donald R
; Detmar, Michael
; Claffey, Kevin P
; TITLE OF INVENTION: Method for inhibiting tumor
; angiogenesis in a living subject
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, Esq.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: Dell PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Microsoft Word version 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/532,310B
; FILING DATE: 22-Mar-2000
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker, Esq.
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: BIS-036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-532-310B-5

Query Match 29.2%; Score 1784; DB 4; Length 1183;
Best Local Similarity 35.2%; Pred. No. 5.7e-163;

Matches 421; Conservative 222; Mismatches 474; Indels 80; Gaps 24;
Qy 13 LVFLTGLCSFNLDEHHRLPFGPEABFYSVLQHVGGQRMWLVGAPWDGSGDRRGD 72
Db 21 LTVILGFCVFNVDVKNMSTFSGPVEDMFGYTVQOYENEEGKWVLIGSPLVGQPKNRTGD 80
Qy 73 VYRCPVCGAHNACAKGHIGDYQLGNSSHPAV-----NMHLGMSLLTDCDGGGFWACAPL 127
Db 81 VYKCPVGRGESLPCVKLDLP-----VNTSIPNVTEVKENMTFG-STLVNPNNGSLACGPL 135
Qy 128 WSRACSSSVFSGICARVDASFQPGSLAPTAQRCPTYMVDVIVLDGNSNIYPPSEVQTF 187
Db 136 YAYRCGLHYTTGLCSDVSPTFQVNVSIAP-VQECSTXLDIVIVLDGNSNIYPPMDSV-TA 193
Qy 188 LRLVLGKLFDPQIOVQLVOYGESPVHWSLGDFTKVEVRAAKNLSRREGRETAKQ 247
Db 194 LNDLLKEMDGPCKXTXGVIKXGENVTHFNLNKYSSTEEVLAAKKIVXRGGRXTWTAL 253
Qy 248 AIWVACTEGFSQSHGGRPEAARLLVVVTDGESHDEELPAALKACAGRVTRYGI AVLGH 307
Db 254 GTDTARKEAFTEARGARGVKVMVIVTDEHDXNHLKKVIGDCEDENIXRFIALLGS 313
Qy 308 YLRQRDPSPFLREIRTIASDPDRFFFNVTDEAALTDIVDALGDRIFGLESGHAENESS 367
Db 314 YNRGNLSTEXFVEEIKSIASEPTEKSFNVSDELALVTIKTLGERIFALEATADXSAAS 373
Qy 368 FGLEMSOIGFSTHRLKDGILFGVMGVADWGVSVLWEGGHRFLPPRMALEDEFPALQNH 427
Db 374 FEMEMSTGFSAHYSQDMVWLGA VGAYDWNGTVVYNQASQIIIPRNTTFNVESTKNEPL 433
Qy 428 AAYLGYSVSSMLRGRRRLFLSGAPRFRHGRKVTAQFKKDGAVRVAQSGOGEIGSYFG 487
Db 434 ASYLGTVNSATASSGDVLYIAGQPRYNHTGQVYIYRM-BEGNIKIQTLSGXIGSYFG 492
Qy 488 SELCPLDTRDGTDLVLLVAAPMFLGPQNKETGRVYVVLVVGQQSLTLTQGTLP-EP--- 543
Db 493 SILTTTIDIKDSNTNILLVGAPMYMGTEKEGGKVYYVAL-NQTRFVQNSLAPMEPIKQ 551
Qy 544 -----PODARFGFANGALPDNLQDGFADVAVGAPLEDHQGALYLYH 585
Db 552 TCCSSRQHNSCTTENKNEPCGARPGTAATAVKDLNLDGFNDIVIGAPLEDHGGAVYIYH 611
Qy 586 GTQGVPRPAPQRIATAASMPHALSYFGSRVDGLDGLDDLDVAVAGQAAILLSRPI 645
Db 612 GSGKTIKEXYAXRIPSGGDKTLKFQGSIHGEMDLNGDGLTDVTICGLGAALFWSRDV 671
Qy 646 VHLTPSLEVTPOAISVVQDRCRRRQBAVCLTAALCFQVTSRTPGRWDHOFYMRFTASLD 705
Db 672 AVVKVTNFPFNKVNIOKKNCHMEGKETVCINATVCFEVLKSKEDITYEADLQYRVTL 731
Qy 706 EWTAGAFAFDGSGQRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRLPVALTFTFALDNTT 765
Db 732 SLRQISRSFSGTQERKVQR-NITVRKSECTKSHFYMLDKHDPQDSVRIITLDF---NLT 786
Qy 766 KP--GPVLNEGSPSIOKLVFPFSKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGGRK 823
Db 787 DPENGVLDDSLPNSVHEYIPFAKDCGCKEKISDLSLHV-----ATTEKDLIIVRSONDK 842
Qy 824 VLVSTTLNKRKENAYNTSLSIIFSRNLHLASTLPQRESPIKVECAAPSAHARLCSVGHVP 883
Db 843 FNVSLTVKNTKDSAYNTRTIVHYSPLNVFSGI-----EAIQKDC--ESNHNITCKVGYF 896
Qy 884 FQTGAKVTFLEPFSCSSLLSQVFGKLTASSDLENGTLQENTAOISAVIQEYPHLLP 943
Db 897 LRRGEMVTFKILFOFNTSYLMENVTIYLSATSDSEEPETLSDNVNVIPIVKYEVLQF 956
Qy 944 SSESTLHRYEHPYGTLP-----VGPGEPEKTLRVQNLCCYVVSGLIISALLPAVAHG 997
Db 957 YSSASEYHISIAANETVPEVINSTEDIGNEINIFYLIRKSGSFPMPPELKLKLSISFNM 1016
Qy 998 GNYFL---SLSOVITNASCIVQNLTEP-----PGPPVHPEELOHTNRLNGSNTQCQV 1048
Db 1017 GYPVLVPTGLSS--SENANCRPHIFEDFPFSINSKKWTSTDLHKKRGTLDCNTCKPATI 1074

```

Qy   1049 RCHLGQLAKTEVS-VGLLRVLVHNEFRRRAKFKSLTVVSTFELGTSEGSVLQLTASRWS 1107
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db   1075 TCNL-----TSSDISXNVSVLILWKTPEIKGVFSLSNLTIRGELRSNAS-LVLSSSNEKR 1129
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy   1108 ESLEVVQT-RPLISILWILGSLVGLLILLALLVCLWKLGCPFAHKKIPIEBEKREE 1163
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db   1130 ELAIQISKDGLPGRVPLMWILLSAFAGLLMLLILALWKIGFF---KRPLKKMKEX 1183
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

RESULT 4
US-09-000-004A-4
; Sequence 4, Application US/09000004A
; Patent No. 6780603
; GENERAL INFORMATION:
; APPLICANT: Tebibary, Photini-Effie
; APPLICANT: Charonis, Aristidis S.
; APPLICANT: Setty, Suman
; APPLICANT: Maurer, Michael
; TITLE OF INVENTION: ANALYSIS OF ALPHA INTEGRINS FOR THE DIAGNOSIS OF DIABETES
; FILE REFERENCE: 600.314USWO
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/001,387
; PRIOR FILING DATE: 1995-07-21
; PRIOR APPLICATION NUMBER: US 60/001,861
; PRIOR FILING DATE: 1995-08-03
; PRIOR APPLICATION NUMBER: US 60/016,700
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: PCT/US96/12067
; PRIOR FILING DATE: 1996-07-19
; NUMBER OF SEQ IDS NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-000-004A-4

Query Match          28.8%; Score 1756; DB 4; Length 1181;
Best Local Similarity 34.3%; Pred.No.3e-160;
Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

Qy   11 LPLVFLTGL-----CSPFNLDHHRLPPGPPEAEFGYSLVQHVGCGQRWMLVGAPWD 63
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db   11 LPULLLVALSQGIILNCCLAYNVGLPEAKIFSGPSSEQGYAVQQFINPKGNWLLVGSFWS 70
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy   64 GPSGDRGDVYRCVPVGAHANPACAKGHL-GDYQIGNSHPHAVNMHLQMSLLETDDGGCFM 122
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db   71 GFPENRMGDVYKCPV-DLSTATCEKLNLTQTSIPNVTEMKNLSGLILTRNNMTGGFL 129
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy   123 ACAPLMNRACGSSVFSSGICARVDASFPQGS LAPTAQRCPETYMDVVIVLDGNSIYPWS 182
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db   130 TCGFLMAQQCGNQYYTTTGVCSDISPDPQLSASFSPATPCPSLIDVVVVWCDSNISYPWD 189
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy   183 EVQTFRLRLVGLKFIDFEIQVGLVQYGESPVHSWSLGDPRTKEEVVRAAKNLSRRSGRE 242
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db   190 AVKNFLEKFVQGLDIGTKTVQGLIQYANPRVVFNLTNYTKREMIVATSQYSQYGGDL 249
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy   243 TKTAQAIIWACTEGFSOSHGRPEARBLLVVVTDGESHGDEELPAALKACEAGRVTRYGI 302
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db   250 TNTPGAQIYARKYAYSAAAGRRSRATKMVMVVTDGSHDGSMLKAVIDQCNDHNDILRFGI 309
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy   303 AVLGHYLRQRDPSPFLREITTIASDPDERFFENVTDAAALTDIVDALGDRIFCLEGSHA 362
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db   310 AVLGYLNRNALDTKNLIKIAIASIPTERYFFNVSDAALLEKAGTIGHQIFSIEGT-V 368
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy   363 ENESSFGLMSQIGFST--HRLKDGIILFMVGAYDWGCVSLW-LEGGHRLPFPPRMALDE 419
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db   369 QGGDNFQMEHMQVGFSDYSSQNDIILMLGAVGAGNGSGTIVOKTSHGHLTFP-----KQA 423
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

Qy   420 PPPALQ--NHAAIYLGYSVSSMLRGRRRLFUSGAPRFRHRKGKVIATPOLKDGAVRVAAQS 477
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::

```

424	Db	FDQILQDRNHSSYLGYSVAA--ISTGSETHPVAGAPRANTGTQIVLVSVNENGNIIVTQAH	482
478	Qy	QGEIQISYSGSELCLPTDRDGTDLVLLVAAPMLGPQNKETGRVTYV-----LVGQOQL	532
483	Db	RGDQISYSGSVLCSVDVDKDTITDVLVAGAPMYMSDLKKEGRVYLFITKKGILQHQHF	542
533	Qy	LITOGTLQPEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEPDHGOHALYLYHGTQSGVR	592
543	Db	--LEG--PEGIENTRFGSAIAALSINMDGPNVDIVGSPLENQNSGAVYVYNGHQQTIR	597
593	Qy	PHPAQRIAAA--SMPHALSYFGBSVDGRDLDDGDDLVDVAVAGQAQAAILSSRPIVHLTP	650
598	Db	TKYSKILGSDGAFRSHLQYFGRLSDGYGLDNGDSITDVISGAPGVQVQLMSQSIADVAI	657
651	Qy	SLEVTFOATSVVORDCRRRQQAECVLTALCFQVTSRTPGRMWDHQVYMRFTASLDSEWTAG	710
658	Db	EASFTEKILVNKNAQ-----IILKCFSAKFR-PTKQNNQVAIVNYNTLDA----	704
711	Qy	ARAAFDGSGQRLSPREL-----RLSVGNV-----TCBQLHPHVLDTSDYLRPVALTVT	758
705	Db	-----DQFSRVTSRGLFKENNERCLQKNMVVNAQSCPEHIIYIQBSPDVVNSLDLRVD	759
759	Qy	FALDNTTKPG--PVINEGSTPSIQKLVPPSKDCGPNECVTDLVLQVNMWDIRSGRAKPFV	816
760	Db	ISLEN---PQTSPALRAYSETAKVFGIPPHKDCGEDGLCISDLVLVDVR-QIPAAQSQPFI	815
817	Qy	VRGGRKVLVSTTLENKENAYNTSLSIIFSRNLHLASLTPQRESPI---KVSC-AAPSA	872
816	Db	VSQNKRGLTFSVTLKKNRGSAYNTGIVDFSENLFPASFS-----LPVDGTETVTCQVAASQ	871
873	Qy	HARLCSVGHVPFQTGAKVTFLEBFSCSLSSQVFGKLTASSDSLRNGTLOQNTAQT	932
872	Db	KSVACDVGVPALKREQVTFITNFDNLQNLQNAQSLSPQALSESQENKA--DNLVNLK	939
933	Qy	AYIQYEPHLLFSSSESTLHRYVHPYGTLP-----VGPGEFXTTLRVQNLGCYVVSGLI	986
930	Db	IPLYDAETHLTRSTINIFYETISDGNVPFSIVHSPEDVGPKFIFSLKV-TTGSVPVSMAT	988
987	Qy	ISALLPAVAGGNYFLSLQVITNNASCTVQNUTEPP-----GPPVHPBELQHTNRL	1038
989	Db	VIIHIQYTKKNPLMYLTGVQTDKAGDISCNADINPLKIQGTSSSVSFKSENFRHTKEL	1048
1039	Qy	NGSNTQCQVYRCHLGQAKGTEVSUGLLRLVHNEFFRRRAKFKSLTVVSTPELGTBEGSVL	1098
1049	Db	NCRTASSCNVTCWLKDVMHKGIFYVNVTRIWNGTFTASSTFQVQLTAAAEINTYNPEIY	1108
1099	Qy	QLTEASRWSESLEUVQVTRPILISLMI-----LIGSVLGGLLLLLALLVFLCW	1145
1109	Db	-----VIEDNTVTIPLIMKPDKEAEVPTGTVIIGSIAGIILLALLVAILW	1154
1146	Qy	KLGFPAHK-----KIPEE	1158
1155	Db	KLGFPRKRYEKWTKNPDE	1172

RESULT 5
 US-09-949-016-6189
 ; Sequence 6189, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ IDS NOS: 207012

RESULT 5

US-09-94

Section

Deborah

Patent:

; GENERA.

APPLI

TITLE :

TITLE :

FILE #

• CITRUS FRUIT

COMING

CUKRE;
DOSTO

;
PRIOR

;
PRIOR

PRIOR

; **PRIOR**

PRIOR

PRTOR

NOTES

NUMBER :

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6189

; LENGTH: 1181

; TYPE: PR

; ORGANISM: Human

US-09-949-016-6189

Query Match 28.8%; Score 1756; DB 4; Length 1181;

Best Local Similarity 34.3%; Pred. No. 3e-160;

Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

```
Qy 11 LPLVFLTGL-----CSPFNLDHHPRLFGPPPEAFSGYVLQHVGGQRMVLVGPWD 63
Db 11 LPLLLVLSQGLNCLAYNVGLPEAKIFSGPSEQFGYAVQVQFINPKGNWLLVGPWS 70
Qy 64 GPSGDRGDVYRCPVGGAHNAPCAKHL-GDYQGNSSHPAVNMHLGSLLETDCGGFM 122
Db 71 GFPENRMGDVYKCPV-DLSTATCEKLNQTSIPNVTMKTNNMNLGLILTRNMTGGFL 129
Qy 123 ACAPLWSRACGSSVFSFGICARVDASFPQGLSLAPTAQRCPTVMYDVLGNSIYPS 182
Db 130 TCGPLWAQCCGNYTITGVCSDISPDFQLSASFATQPCPSLDIVVVCDESNIYPWD 189
Qy 183 EVQTFLLRLVGLKFLDPEIQVGLVQYGESVPVHWSLGDFTKEEVRAAKNLSRREG 242
Db 190 AVKNFLEKFGQLDIPGKTQVGLIQYANPRVFNLTNTYKTEEMIVATSQTSQYGGDL 249
Qy 243 TKTAQAIWVACTEGFSQSHGGRPEARLLVVVTGESHGDEELPAALKACBAGRVTYGI 302
Db 250 TMTFGAIQYARKYASAAAGRRSATKVMVVVTGESHGDSMLKXAVIDQCNDHNLRFGI 309
Qy 303 AVLGHYLRQRDPSPFLREITIASDPERFFNVVTDEAALTDIVDALGDRIFGLESHA 362
Db 310 AVLGYLARNALDXTNLKAIKAIASIPTEYFFNVSDAALLKXAGTLEQIFSIETG-V 368
Qy 363 ENESFGLNSQIGFST--HRLKDGILFGMYGADWGSVLW-LEGHRLPPRMALEDE 419
Db 369 QGGNFQWMSQVGSADYSQNDILMLGAVGFWGSGTIVKTSNHLIFP-----KQA 423
Qy 420 FPPALQ--NHAAYLGYSVSMRLGRRRLPLSGAPRFRHRKGVIAFOLKKGAVRVAQSL 477
Db 424 FDQILQDRNHSYLGYSVAA-ISTGESTHFVAGAPRANYTQGIYLVSWENGNITVIAH 482
Qy 478 QGEQIGSYFSGELCPDTRDGTDLVLLVAPMFLGPONKETGRVYV-----LVGQSL 532
Db 483 RGDQIGSYFSGVLCSDVDKDTITDVLVAGAPMTMSDLKKEGRVYLFITKKGILGHQF 542
Qy 533 LTLQGLTLOPEPPQDARFGMAGLPDLNODGFADVAVGAPLEDHOGALYLYHGTQSGVR 592
Db 543 --LEG---PEGIENTRFGSAALSDINMGDFNIVVGSPLNQNSGAVIYNGHQGTIR 597
Qy 593 PHPAQRIAAA--SMPHALSIFGRSDGRDLDDGDLVDVAVGAQGAAILLSRRPIVHLTP 650
Db 598 TKYSQKILGSDGAFRSHLOYFGRSLDGYGLNGDSITDVSIGAGQVVLWSQSIADVAI 657
Qy 651 SLEVTPOAISVQDRCRRRQGEAVCLTAALCFQVTSRTPGWDHOFYMRFTASLDWTAG 710
Db 658 EASPTPEKITLVNNAQ-----IILKCFSAKFR-PTKQNNQVAIVNITLDA----- 704
Qy 711 ARAAFDGSGBRLPRRL-----RLSVGNV-----TCEQLHFHVLDTSDYLRPVALT 758
Db 705 -----DGFSSRVTSRGLPKENNERCLQKNVMVNAQSCPEHIIYIQEPSDVVNSLDURVD 759
Qy 759 FALDNTTKPG--PVLNKGSPSISQKLVFSDKCPDNECVTDVLQVNMIDIRGRKAPFV 816
Db 760 ISLEN---PGTSPALEAYSETAKVFSIPFKDCGDEGLCISDLVLDVR-QIPAAQEQPFI 815
Qy 817 VRGGRKVLVSTTLNKRKENAYNTLSIIFSRNLHSLATPQRESPI---KVEC-APPSA 872
Db 816 VSNQKRLTFSTVLKKNKESAYNTGIVVDFSENLFASFSS---LPVDGTETVTCQVAASQ 871
Qy 873 HARLCSGVHPVQTCAGKATFLELFEFFSCSLLSQVFGKLTASSDSLENGTLQENTACTS 932
```

```
Db 872 KSVACDVGPALKREQQVTFITNFDNLQNLQNASLSFOALSSESQENKA--DNLVNLK 929
Qy 933 AYIQYEPHLLPSSESTLHYEVHPYGTLP-----VGPGEFKTTLVQNLGCVVSGLI 986
Db 930 IPLLDAEHLTRSTNINFYEISDGNVPISVHSFEDVGPKFIFSLKV-TTGSVPVSNAT 988
Qy 987 ISALLPAVAGGNYFLSLSOVITNNASCIQNLTEPP-----GPPVHPEELQHTNRL 1038
Db 989 VIHIPOYTEKPLMYLTGVQTDKAGDISCNADINPLKIGOTSSSVSFKSENFRTKEL 1048
Qy 1039 NGSNTQCVVYRCHLQAKGTEVSGLLRLVHNEFFRRAKPKSLTVVSTFELGTEGSVL 1098
Db 1049 NCRTASCSNVTCLWKDVHMKGEYFVNVTVTRWNGTFASSTFQTVQLTAAABINTVPEIY 1108
Qy 1099 OLTEASRWSESLLEVQVTRPILISLWI-----LIGSVGLLILLALLVFCILW 1145
Db 1109 -----VIEDNTVTIPLIMTKPDEKAEVPTGVIIGSIAGILLALLAILW 1154
Qy 1146 KLGFPAHK-----KIPPE 1158
Db 1155 KLGFKKRYEKMYKNPDE 1172

RESULT 6
US-09-949-016-10747
; Sequence 10747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10747
; LENGTH: 1195
; TYPE: PR
; ORGANISM: Human
US-09-949-016-10747

Query Match 28.8%; Score 1756; DB 4; Length 1195;
Best Local Similarity 34.3%; Pred. No. 3e-160;
Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

Qy 11 LPLVFLTGL-----CSPFNLDHHPRLFGPPPEAFSGYVLQHVGGQRMVLVGPWD 63
Db 25 LPLLLVLSQGLNCLAYNVGLPEAKIFSGPSEQFGYAVQVQFINPKGNWLLVGPWS 84
Qy 64 GPSGDRGDVYRCPVGGAHNAPCAKHL-GDYQGNSSHPAVNMHLGSLLETDCGGFM 122
Db 85 GFPENRMGDVYKCPV-DLSTATCEKLNQTSIPNVTMKTNNMNLGLILTRNMTGGFL 143
Qy 123 ACAPLWSRACGSSVFSFGICARVDASFPQGLSLAPTAQRCPTVMYDVLGNSIYPS 182
Db 144 TCGPLWAQCCGNYTITGVCSDISPDFQLSASFATQPCPSLDIVVVCDESNIYPWD 203
Qy 183 EVQTFLLRLVGLKFLDPEIQVGLVQYGESVPVHWSLGDFTKEEVRAAKNLSRREG 242
Db 204 AVKNFLEKFGQLDIPGKTQVGLIQYANPRVFNLTNTYKTEEMIVATSQTSQYGGDL 263
Qy 243 TKTAQAIWVACTEGFSQSHGGRPEARLLVVVTGESHGDEELPAALKACBAGRVTYGI 302
Db 264 TMTFGAIQYARKYASAAAGRRSATKVMVVVTGESHGDSMLKXAVIDQCNDHNLRFGI 323
Qy 303 AVLGHYLRQRDPSPFLREITIASDPERFFNVVTDEAALTDIVDALGDRIFGLESHA 362
```

```
Db 324 AVGLYLRNALDTKNLIKAKATASIPTEYFFNVSDAALKEKAGTLGQIFSIECT-V 382
Qy 363 ENBSFGLMSOIGFST--HRLKDGILFGVGVADMGGSVLM-LEGHRLFPFPRMALEDE 419
Db 383 QGDNFQWMSQVGSFADYSQNDILMLGAVGAFWGSSTIVQKTSHGHLIFP-----KQA 437
Qy 420 FFPALQ--NHAAYLYGVSVMMLRGRRLLFLSGAPRFRHKGKVIQFOLKDKDGVAVRVAQSL 477
Db 438 FDQLQDRNHSSYLYGSVAA-ISTGBSTHFVAGAPRANYTGOVLVSVNENGNIIVLQAH 496
Qy 478 QGEOIGSYFGSELCPDLTDRDGTDLVLLVAPMFLGPONKETGRVYVY-----LVGQOQL 532
Db 497 RGDQIGSYFGSVLCSVDVDKDTITDVLVAGAPMTMSDLKKEBGRVYLFTEKGLGQHOF 556
Qy 533 LTQGLTQPPPPQDARFGFAGVCLPDINQDGFADVAVGAPLEDHGHQCALVLYHCTQSGVR 592
Db 557 --LEG---PGIENTRFGSAIAALSDINMGDFNDVIIVGSPLENGNSGAVIYNGHQTIR 611
Qy 593 PHPAQRIAAA--SMPHALSFGRSVDGRLLDLGDGLVDVAVAGQAAILSSRPVHLTP 650
Db 612 TKYSQKILSDGAPRSHLQFGRSLDGYGLNGDSITDVISGAFQGVVQLWSQSIADVAI 671
Qy 651 SLEVTPOAISVVQDRCRRRQGEAVCLTAALCFQVTSRTPGRWDHQFVWRFTASLDEWTAG 710
Db 672 EASFTPEKITLVNKAQ-----IILKLCFSAPFR-PTKQNNQVAIVYNIITLDA---- 718
Qy 711 ARAAFDGSGORLSPRL--RLSVGNV-----TCEQLHFLVLDTSVLRPVALTVT 758
Db 719 -----DGFSSRVTSRGLFKENNERCLQNMVYVNAQSCPBEHIIYIQSPDVNSLDLRVD 773
Qy 759 FALDNTTKPG--PVNLGSPSTSIQKLVFPGKDCPDNECVTDLVLVQNMIDIRSRKAPFV 816
Db 774 ISLEN---PGTSPALEAYSETAKVPSIPFHKDCGEDGLCISDLVLDVR-QIPAAQEQPFI 829
Qy 817 VRGRRKVLVSTTLNENKENAYNTLSIIFSRNLHLASLTPORESPI---KVEC-AAPSA 872
Db 830 VSNQKRLTFSVTLKKNRESAYNTGIIVDFSENLFASFS----LPVDGTEVTCQVAASQ 885
Qy 873 HARLCSVGHVPVFGAKVTLLEPFSCSLLSQVFGKLTASSDSLSRNGTLQENTAQTS 932
Db 886 KSAACDVGYPALKREQQVTFINFDNLQNLQNASLSFQALSSESQENKA--DNLVNLK 943
Qy 933 AYIQEYPHLLFSSBSTLHREVEHPYGLP-----VGPGEFKTTLRVQNLGCVVWSGLI 986
Db 944 IPLYDAEIHLTRSTNINFYISSDGNVPSIVHSEFEDVGPKEFISLKV-TTGSVPVSMAT 1002
Qy 987 ISALLPAVAGGNVFLSLSQVITNNASCTIVONLTTEPP-----GPPVHPPELOHTNRL 1038
Db 1003 VIIHIPQVTKENPLMYLTGVQTDKAGDISCNADINPLKIGQTSSSVSFKSENFRHTKEL 1062
Qy 1039 NGSNTQCOVVRCHLQOLAKGTEVSGLLLRVHNEFFRRRAKFKSLTVVSTPELGTTEGSVL 1098
Db 1063 NCRFASCSNVTCWLKDVHMKGEYFVNTVTRWNGTFASSTFTQVLTAAABINTPEIY 1122
Qy 1099 QLTEASRWSLSLLEVVQTRPILISLWI-----LIGSVLGLLILALLVFLCW 1145
Db 1123 -----VIEDNTVTIPLMIMKPEKAEVPTGVIGSIIGIALLLALLVALIW 1168
Qy 1146 KLGPFAHK-----KIPBE 1158
Db 1169 KLGFPRKYEKMTKNPDE 1186
```

RESULT 7

US-09-532-310B-6

Sequence 6, Application US/09532310B

Patent No. 6596276

GENERAL INFORMATION:

APPLICANT: Senger, Donald R

Detmar, Michael

Claffey, Kevin P

TITLE OF INVENTION: Method for inhibiting tumor

angiogenesis in a living subject

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: David Prashker, Esq.
STREET: P.O. Box 5387
CITY: Magnolia
STATE: Massachusetts
COUNTRY: USA
ZIP: 01930
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
COMPUTER: Dell PC
OPERATING SYSTEM: MS DOS
SOFTWARE: Microsoft Word version 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/532,310B
FILING DATE: 22-Mar-2000
CLASSIFICATION: Unknown
ATTORNEY/AGENT INFORMATION:
NAME: David Prashker, Esq.
REGISTRATION NUMBER: 29,693
REFERENCE/DOCKET NUMBER: BIS-036
TELECOMMUNICATION INFORMATION:
TELEPHONE: (978) 525-3794
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1183 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-532-310B-6

Query Match 28.3%; Score 1728; DB 4; Length 1183;

Best Local Similarity 33.7%; Pred. No. 1.5e-157;

Matches 412; Conservative 229; Mismatches 459; Indels 124; Gaps 32;

```
Qy 11 LPLVFLTGL-----CSPENLDEHHPRLPPGPGPEAFEGYSVLQHVGGGQWMLVGPWD 63
Db 11 LPLVLLALSQGLINCLAYNLVGLPEAKIPSGSSQFGYAVQOQFINPKGNLLLVGSPWS 70
Qy 64 GPSGDRRGDGYRCPVCGAHNAPCAKGLH-GDYQLGNSSHPAVNMHLGMSLLETDDGGGPM 122
Db 71 GPFENRMGDVYKCPV-DLSTATCEKLNLTQSTSPNVTEMKTNMSLGLILTRNMGTCGFL 129
Qy 123 ACAPLWSRACSSVFSGGICARVDASFQPGSLAPTAQRCPTVMDVIVLDGNSIYFWS 182
Db 130 TCGPLMAQQCGNYTTGVCSDISPDFXKSASFSPATXPCPSLIDVVVVCDESNIYFWD 189
Qy 183 EVOTFLRLVLGKLFIDPEQIQVGLVOYGESPVHWSLGDFTKBEVVRAAKNLSRRREGRE 242
Db 190 AVKNLEKFKVQGLDIDGTTKQVGLIQYANNPRVFNLTYYTKTBEMIVATSTQSYGGDL 249
Qy 243 TKTAQAIMVACTBFGSQSHGGRPEAARLLVVYVTDGESHDEELPAALKACEAGRVTYGI 302
Db 250 TMTFAGIQYARKYVSAAGSGRSATKVMVYVTDGESHGDSMLKAVIDQCNDHNLRFGI 309
Qy 303 AVLGHYLRORDPS-SFLREIRIIVASDPDRPFNVYDDEAALTDIVDALGDRIFGLEGSH 361
Db 310 AVLGYNLRNALDTKNMLIKEIKAISIPTEYFNFVSDEAALLKAGTLGQIFSIEGT- 368
Qy 362 AENESSFGLMSOIGFST--HRLKDGILFGVGVADMGGSVLM-LEGHRLFPFPRM--AL 416
Db 369 VVGDNFYMMSQVGSFADYSQNDILMLGAVGAFWGSSTIVQKTSHGHLIFPKAQFQOI 428
Qy 417 EDEFPFPAQNHAAAYLYGVSVMMLRGRRLLFLSGAPRFRHKGKVIQFOLKDKDGVAVRVAQS 476
Db 429 QD-----RNHSSYLYGSVAA-ISTGBSTHFVAGAPRANYTGOVLVSVNENGNIIVQA 481
Qy 477 LQGEQIGSYFGSELCPDLTDRDGTDLVLLVAPMFLGPONKETGRVYVY-----LVGQOQS 531
Db 482 HRGDQIGSYFGSVLCSVDVDKDTITDVLVAGAPMTMSDLKKEBGRVYLFTEKGLGQHOF 541
```

QY 532 LLTLQGLQPEPPQDARFGAMGALPDLNQDFADVAVGAPLEDHOGALYLYHGTSQGV 591
Db 542 F--LEG---PEGIENTFGSAIALSINMDGDFNDVIVGSPLENQNSGAVIYNGHGTI 596
QY 592 RPHPAQRIAAA--SMPHALSIFYGRSVDGRDLDDGLVDVAVAGAGAAILLSSRPVHLT 649
Db 597 RTKYSQKILSGDGAFAHSLQYFGSLDGYGLDGLSDITDVISIGAFGVQVQLWSQSIADVA 656
QY 650 PSELTPOALISVQDRCRRRQGBACVITAAALCFQVTSRTPGRWDPHQFYMRTASLDWETA 709
Db 657 IEASFTPEKITLVNKAQ-----IILKLCFSAKFR-PTKONQVAIVNITLDA--- 704
QY 710 GARAADGSGORLSPRL-----RLSVGNV-----TCEQLHFLVLDTSYLRPV-ALT 756
Db 705 -----DGFSSRVTSRGLFKENNERCLOKXNVVNOAQSCPEHIYIQEPSDVNSLDDL 758
QY 757 VTFALDNTTPKG--PVLNKGSPGTSIQKLVPPSKDCGPDNECVTDVLQVNNMDIRGSRKAP 814
Db 759 VDLSLEN---PGTSPALEANSETAKVPSIPPHKDCGEGKISDLVLQ-DVRIAPAAQEP 814
QY 815 FVVRGRRKVLVSTLENKRNAYNTSLISIFSRNLHLASLTPORESFI---KYEC-AAP 870
Db 815 FIVSNONKRLTFSVTLKNKRESAYNTGIVVDFSENLFASFS---LPVDGTEVTCQVAA 870
QY 871 SAHARLCSVGHVPQTCAKVTFLLEPFESCSLLSQVFGKLTASSDSLSRNGTLQENTAQ 930
Db 871 SQKSVACDVGYPALKRQEQVTFITFDNPLNQLNQASLSFQALSSESQENKA--DNLVN 928
QY 931 TSAVIOVEPHLLFSSESTLHRYEHPVGTLP-----VGPCEPKTTLRVQNGLCYVVS 984
Db 929 LKIFLLDAEHLTRSNINFIYESSDGNVPSIVHSFEDVGPKFIFSLKV--TTSQSVPM 987
QY 985 LIISALLPAVAGNGNYFLSLSQVITNNASCIVQNLTEPP-----GPPVHPELOHTN 1036
Db 988 ATVIIHIPQVTKENPLMYLTGVQTKAGDCCNADINPLKIGOTSSSVFSKSENFRHTK 1047
QY 1037 RLNGSNTQCVVRCHLQLAGTEVSGCLLVLNVEFRRAKFKSLVTVSTFELGTSEGS 1096
Db 1048 ELNCRTASCNVTCWLKDVHMKGEYFVNVTRIWNGTFASTFTQVQLTAAAEINTYNPE 1107
QY 1097 VLQUTEASRSESLVVOVTRPILISLMI-----LIGSLGGLLMLALLVFC 1143
Db 1108 IY-----VIENTVTIPLMKPDEKAEVPTGVIIGSIAGILLALLVAI 1153
QY 1144 LWKLGFFAHHKIPBEKREEKLEQ 1167
Db 1154 LWKLGFFKRYEKMTKNPDEIDE 1177

RESULT 8

US-08-173-497-2
; Sequence 2, Application US/08173497
; Patent No. 5437958
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van Der Vieren, Monica
; TITLE OF INVENTION: No. 5437958el Human 2 Integrin Alpha
; TITLE OF INVENTION: Subunit
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 S. Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/173,497
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5437958and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/31363

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-173-497-2

Query Match 18.5%; Score 1131.5; DB 1; Length 1161;

Best local Similarity 29.4%; Pred. No. 1e-99;

Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDEHHPRLPFGPPEAEFGVSLQHVGGORWMLCAPWDPGSDRRGVDVYRCPVCGAH 82

Db 17 FNLDVREPTIFQ-EDAGGFGQSVVQF--GGSRLVVGAPLEVVVAANTGRLYDC---AAA 69

QY 83 NAFCAKGHLDYQLGNSHPAVNMHGLMSLLETDGDFMACAPLWGRACGSSVFSFGIC 142

Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASNGSRLLACGPTLHRVCGENSYSKGC 123

QY 143 ARVDASFQPGSLAPTAQRCPT-TYMDVVIVLDGNSI--YPMSEVQTFRLRLVKGLFIDP 199

Db 124 LLLGSRWEIITQVTPDAPTECPHOEMDIVFLIDGSGSIDQNDFNQMGKFVQAVMQQ--FEG 181

QY 200 EQIQVGLVQVSGSPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKTAQAMVACTEGFSQ 259

Db 182 TDTLFAIMQVSNLKIHTFTQPTSPSQSLVDPIVQLKGL-TFTATGILTVVTQLFHH 240

QY 260 SHGRPEAARLLVVTVDGESH-DGEELPAALAKACEAGRVTYRGIYAVLGHYLRRLRDPSSF 318

Db 241 KNGARKSAKILVITDQKQKPLEYSDVIPOAEKAGIIRYAIQV-GHAF---QGFTA- 295

QY 319 LRIRITIASDPDRFFNFVTDAAALTDIVDALGRIFGLEGSHAENESSFLEMSQIGFS 378

Db 296 ROELNTISSAPPQDHFVKVDNFAALGSIQKQLOEKIYAVEGTQSRASSSFQHEMSQEGFS 355

QY 379 THRLKDGILFGMVGAYDMGGSVLWLEGGHRLFPFMALEDEFFPALONHA-----AY 430

Db 356 TALTMGDLFLGAVGFSW-----SGGAFLYPPNMS-----PTFINMQENVDMRDSY 402

QY 431 LGYSVSMMLRGRRFLSGAPRFRHRGKVIQAF-QLKKDGAVRVAQSLQGEIQIGSYFGSE 489

Db 403 LGYSTELALWKGVQNLVL-CAPRYQHTGKAVITQVSRQW--RKAEVTTGTQIGSYFGAS 459

QY 490 LCPDLTDRDGTTLVLLVAAPMFLGPQNKETGRVYVYLV--GQOSSLTLQGTLOPEPPQD- 546

Db 460 LCSVDVDSGSDTLILITGAPHYV--EOTRGQGVSVCLPRGQVQVQWCDVAVLGEQGHWP 517

QY 547 ARGFAMGALPDLNODGFADVAVGAPLEDHOGALYLYHG-TQGVTPHQAORTAAASMP 605

Db 518 GRFGAALTVLIGDVNEDKLIIDVAIGAPGEQENRGAVLFGHASESGISPSHSQRASSQLS 577

QY 606 HALSYFGRSVDGRDLDDGLVDVAVAGAGAAILLSSRPVHLTPSLEVPQAISVVQRD 665

Db 578 PRLOYFGQALSGGQDLTQDGLMDLAVGARGQVLLRLSLPVLKVCVAMRFPSEVAKAVYR 637

QY 666 C-----RRRQGEAVCLTAALCFQVTSRTPGRWDPHQFYMRTASLDWETAGAPAFDGS 718

Db 638 CWEKPSALEAGDATVCLTIQ-----KSLDQLDGIQSSVRFDLALDPLGRILTSRAIFNET 692

QY 719 GQRLSPRLRLSVGNVTCEQLHFLVLD-TSDYLRPVALTVTFALDNTTKG-----PVLN 772

Db 693 KNPTLTRKTLGLG-IHCETLKLLPDCVEDVWSPIILHLNLSLVRBPFPSPQNLRPVLA 751
Qy 773 EGSPTSQKLVPRKDCGPNCEVTDLVLQNMDIRSKAPFVVRGRKVLVSTTLEN 832
Db 752 VGSQDLFTASLPFRKNGCQDLCEGD--LGVTLFSGLQT---LTVGSSLELNVITVWN 806
Qy 833 RKENAYNTSLIIPSRNLHLASLTPQRES- ----IKVEC-AAPSAHARL----CSVGHPV 883
Db 807 AGEDSYGVTVSLYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLSSRCSVNHPI 866
Qy 884 FQCAKVTFLLEFPFSCSSLLSQVFGKLTASSDLERNGLQENTATQSAIYQEPHLF 943
Db 867 FHEGNGTFIVTFDVSFKATLGDRM-LMRASSENKASSKATFQLELFPKYAVTMI 925
Qy 944 SSESTLHRYEHPYGTLPVGPGRFETTLRVNLCVGVVSGLLIISALLPAVAHGNVFLS 1003
Db 926 SRQESTKY--FNPATSDKKMEABHRYRVNLSQDRLA-ISINFVPLVLLNGAVVMDV 982
Qy 1004 LSQVITNNAACIVQNLTEPPGPPVHPELOHTNR---LNGSNTQCQVVRCHLQOLAKGTE 1060
Db 983 VMEAPSQSLPCVSR-----KPPQHSDFLTQISRSPMLDCSIADCLQFRCDFSPFSVQEE 1037
Qy 1061 VSVGLLRLVNEFPFRRAKFKSLTVSTFELGTEGVSQVLTQTEASRWSESLEL-VVQTRPI 1119
Db 1038 LDFTLKNLFGWVRETLOKKVLVSVAEITFDTSVTSQLPQGFAPFRAQMEVLEBDEV 1097
Qy 1120 LISLWILIGSVLGLLALLLALLVCLWKLGPFAHKKIPEREKREE 1163
Db 1098 YNAIPIWGSGVALLLALITATLYKLGFPRHYKMLEDKPED 1142

RESULT 9
US-08-286-889-2
; Sequence 2, Application US/08286889
; Patent No. 5470953
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Mich
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELETYPE: 312-474-6300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-286-889-2
Query Match 18.5%; Score 1131.5; DB 1; Length 1161;
Best Local Similarity 29.4%; Pred. No. 1e-99;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;
Qy 23 FNLDHHPRLFPQPPBAEFYSVLQHVGGQRMMLYGAPMDGSPGDRRGDVVRCVPGGAH 82
Db 17 FNLDVEEPTIFQ-EDAGGFGQSVVQF--GSSR-LVVVGAPLEVVAANQTGLDYDC---AAA 69
Qy 83 NAPCAKHGLDYOLGNSHPAVNMHLGMSLLETDDGGGFMACAPLMSRACGSSVFPSSGIC 142
Db 70 TGMQCPIPL-----HIRPEAVNMSLGLTLAASNGSRLLACQPTLHRCGENSYSGKSC 123
Qy 143 ARVDASFPQGS LAPTAQSCP-TYMDVVLVLDGNSI--YPMSEVQTEFLRLVGLFIDP 199
Db 124 LLGSRWEIITQVPTDAPTECPHOEMDIVFLIDGSGSIDQNDPFQMGKGFQVAVMGQ--FEG 181
Qy 200 EQIQVGLVQYGESPVHWSLGDPTKEEVVRAAKNLSRREGRETKTAQAIMVACTEGFSQ 259
Db 182 TDTLPALMQYSNLLKTHFTPTQERTSPSQSLVDPIVLQKGL-TFTATGLITVVTQLFHH 240
Qy 260 SHGSRPEAARLLVVTGESH-DGEBLPAALKACEAGRVTRYGIIVLGHVLRQRDPSSF 318
Db 241 KNGARKSAKKILVITDGQYKDPLEYSVDVPOAEKAGIIRYAIGV-GHAF---QGPTA- 295
Qy 319 LRIRTIASDPDERFPFNVTDAAALTDIVDALGDRIFGLEGSHAENESSFGLSEMSQIGFS 378
Db 296 RQELNTISSAPPQDHVFKVDNFAALGSIQKQLOEKIYAVEGTOSRASSSPQHEMSQEGFS 355
Qy 379 THRLKDGILFGMVGAYDWGGSVLMLEGHRLFPFRMALEDEFPPLQNHIA-----AY 430
Db 356 TALTMGDLFLGAVGSFSW-----SGGAPLYPNMS-----PTFINMSQENVMDRDSY 402
Qy 431 LGYSVSMLLRGRRRLFLSGAPFRHRGKVIAF-QLKXGAVVAOSLOGEQIGSVFGSE 489
Db 403 LGYSTELALMKGVQNLVL-GAPRYQHTKAVITQVSROW--RKAEBVTGTQIGSVFGAS 459
Qy 490 LCPDTRDGTDTDLVAAAMPFLGPQNKETGRVYVVLV--GQOSLLTLQGTLOPEPQD- 546
Db 460 LCSVDVDSGSDTLILIGAPHYY--BQTRGGQVSVCLPRQGVQWQCDAVLGEGQHPW 517
Qy 547 ARFGFAMGALPDNLQDGFADVAVGAPLEDDHQCGLYLYHG-TQSGVRPAPQAIAAASMP 605
Db 518 GRFGAALTVLGDVNEKLDLVAIGAPEGQENRGAVLPHGASGSGISPHSQRIASSQLS 577
Qy 606 HALSYFGRSVDRLDLDGDDLDVAVAGAAAILLSRPIVHLTPSLEVTPOAISVVQRD 665
Db 578 PRLQYFQALSGQDLTQDGLMDLAVGARGQVLLSLPVLKVGAVMRFSPVEVAKAVYR 637
Qy 666 C-----RRGQEAVALTAALCFQVTSRTPGRWHDQFYMRTFASLDWETAGARAAPDGS 718
Db 638 CWEEKPSALEAGDATVCLTIQ-----KSLDQGLDIQSSVRFDLALDPGLTSLRAIFNET 692
Qy 719 GQRLSPRRRLSVGNVTCEQLHFPVLD-TSDYLRLPVALTVTFALDNTTKPG-----PVLN 772
Db 693 KNPTLTRKTLGLG-IHCETLKLLPDCVEDVWSPIILHLNLSLVRBPFPSPQNLRPVLA 751
Qy 773 EGSPTSQKLVPRKDCGPNCEVTDLVLQNMDIRSKAPFVVRGRKVLVSTTLEN 832
Db 752 VGSQDLFTASLPFRKNGCQDLCEGD--LGVTLFSGLQT---LTVGSSLELNVITVWN 806
Qy 833 RKENAYNTSLIIPSRNLHLASLTPQRES- ----IKVEC-AAPSAHARL----CSVGHPV 883
Db 807 AGEDSYGVTVSLYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLSSRCSVNHPI 866
Qy 884 FQCAKVTFLLEFPFSCSSLLSQVFGKLTASSDLERNGLQENTATQSAIYQEPHLF 943
Db 867 FHEGNGTFIVTFDVSFKATLGDRM-LMRASSENKASSKATFQLELFPKYAVTMI 925
Qy 944 SSESTLHRYEHPYGTLPVGPGRFETTLRVNLCVGVVSGLLIISALLPAVAHGNVFLS 1003
Db 926 SRQESTKY--FNPATSDKKMEABHRYRVNLSQDRLA-ISINFVPLVLLNGAVVMDV 982

Db 1098 YNAIPIIMGSSVGGALLLALITATLYKLGFFKRYKEMLEDKPED 1142

RESULT 11

US-08-362-652-2

; Sequence 2, Application US/08362652

; Patent No. 5766850

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sear Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/362.652

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/173,497

FILING DATE: 23-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,889

FILING DATE: 5-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/32391

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-362-652-2

Query Match 18.5%; Score 1131.5; DB 1; Length 1161;

Best Local Similarity 29.4%; Pred. No. 1e-99;

Matches 346; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

Qy 23 FNLDHHPRLPPGPEAEFGYSLVQHGQGMWLVGAPMDGPGDRRGDYRCVGGAH 82

Db 17 FNLDVEEPTIFQ-SDAGGFGGVQF--GGSR-LVVGAPLEVVAANTGRLYDC--AAA 69

Qy 83 NAPCAKHLGDYQLGNSSHAPVANNHGLMSLLETDCGGFMACPLMSRACSSVSSGIC 142

Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASTNGSRLLACGPTLHRVCGENSYSKGC 123

Qy 143 ARVDASFPQGSGLAPTAQRCP-TYMDVVIVLDGNSI--YPMSEVQTFRLRLVCKLFDIP 199

Db 124 LLLGSRHEIITVPDAPTECHQEMDIVFLIDGSGSIDQDNFNKMGFVQVMQ--FEG 181

Qy 200 EQIQVLVQYGESPVHMSLGDFTKKEVWRAAKNLSRREGRETAKTAQAINVACTEGFSQ 259

Db 182 TDTLPALMQYNSLKKIHFTFTQFRTPSQSSQSLVDPPIVQLKGL-TFTATGILTVVTQLFHH 240

Qy 260 SHGRPERARLLVVVTDGESH-DEEELPAALKACEAGRVTRYGIYAVLGHYLRDRDPSSP 318

Db 241 KNGARKGAKKILIVITDGQYKDPLEYSDDVIPQAEKAGIIRYAGV-GHAF--QGPTA- 295

Qy 319 LRERTIASDPDRFFNFVTDEAALTDIVDALGDRIFGLEGSHAENESSFGLEMSQIGFS 378

Db 296 RQELNTISSAPPQDHPVKVDNFAALGSIQKQLOEKIYAVEGTQSRASSSFQHEMSQEGFS 355

Qy 379 THRLKDGILFGMWGAYDWGGSVLMLEGGHRLPPRMALEDEFPFALQNH-AY 430

Db 356 TALTMGDLFLGAVGSPSW-----SGGAFLYPPNMS-----PTFINNSQENVDMDRDSY 402

Qy 431 LGYSVSSMLLRGGRRLFLSGAPRFRHKGVIAP-QLKKGAVRVAQSLQSGVSGSE 489

Db 403 LGYSTELALWKGVQNLVL-GAPRYOHTGKAVFTQVSQW--RKKAETVGTQIGSYFGAS 459

Qy 490 LCPDLTDRDGTTDVLLVAAPMFLGPQNKETGRVYVYL-V--GQOSLLTLQGTLOPEPQD- 546

Db 460 LCSVDVDSGSDTLILIGAPHYY--EOTRGQVSVCPPLRGQVQWQCDVLEGEQHPW 517

Qy 547 ARFGFANGALPDINQGFADVAVGAPLEBHQALYLYHG-TOSGYRPHPAQRIAAASMP 605

Db 518 GREGAALTVLGDVNEKLDLIDVAIGAPGEQENRGAVYLFHGASBSGISPSHSQRIASSQLS 577

Qy 606 HALSYFGRSVDGRLDLDGDDLDVDAVGAQGAAILLSRPVHLTPSLVTPQISVVQRD 665

Db 578 PRLQYFGQALSGGQDLTQDGLMDLAVGARGQVLLLSLPLVKVGVAMRFPSPVEKAVYR 637

Qy 666 C-----RRRQGEAVCLTAALCFQVTSRTPTGRWDHQFYMRFTASLDEMTAGARAAFDGS 718

Db 638 CWEKPSALEAGDATVCLTIQ-----KSSLDQLDGDIQSSVRFDLALDPGLTSAIFNET 692

Qy 719 GQRLSPRLRLSVGNVTCQLHPHVL-D-TSDYLRPVALVTVPALDNTTKPG-----PVLN 772

Db 693 KNPTLTRRKTLGLG-IHCETLKLLPCDVEDVVSPIILHLNLSLREPIPSQNLRLPVL 751

Qy 773 EGSPTSIOQLVPFSKDCGPDNECVTLVLQVNMDIRSKAPFVVGRRKVLVSTLEN 832

Db 752 VGSQDLFTASLPFEKNCQDGLCEGD--LGVTLSPSGLQF--LTVGSSLELNVIVTVN 806

Qy 833 RKENAYNTSLIIFSRNLHSLTPORESF-----IKVEC-AAPSAAHARL-----CSVGHPV 883

Db 807 AGEDSGTVSVLYYPAGLSHRRVSGAQKQPHQSALRLACETVTEDEGLSSRCVSNHPI 866

Qy 884 FOTGAKVTLLFEFSCSSLLSQVFGKLTASSDSLRNGTFLQNTAQTSAIYIEPHLLF 943

Db 867 FHEGSGTFTVTPDVSKATLGDRM-LMRASASSENKASSKATQLELPVKYAVVTMI 925

Qy 944 SSESTLHRYEHPYGTLPVGPGEFTLAVQNLGCYVVSGLLIISALLPAVAGNRYFLS 1003

Db 926 SROEESTKY--FNFATSDKKMKAEHRYRVNLSQORDLA-ISINFVPVLLNGVAVMDV 982

Qy 1004 LSQVITNNASCIVONLTETPPGPPVHPEELQHTNR---LNGSNTQCQVVRCHLQOLAKGTE 1060

Db 983 VMEAPSQSLPCVSE-----KPPQHSDFLTQISRPMLDCSIADCLQPCDVPSPVQEE 1037

Qy 1061 VSVGLLRVHNEFFRRAKFKSLTVVSTFELGTBEGSVLQLTASRWSESLLE-VVQTRPI 1119

Db 1038 LDFTLKNLSFGWVRETLOKKLVWSVAEITFTSVYSQLPGQEAQFMAQMEMVLEDEV 1097

Qy 1120 LISLWILGSLVGLHLLALLVFLCKLGFPA-AHKKIPEEKEE 1163

Db 1098 YNAIPIIMGSSVGGALLLALITATLYKLGFFKRYKEMLEDKPED 1142

RESULT 12

US-08-605-672-2

; Sequence 2, Application US/08605672

; Patent No. 5817515

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun


```
/ APPLICATION NUMBER: US/08/482,293A
/ FILING DATE:
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/173,497
/ FILING DATE: 23-DEC-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/286,889
/ FILING DATE: 5-AUG-1994
/ APPLICATION DATA:
/ APPLICATION NUMBER: US 08/362,652
/ FILING DATE: 21-DEC-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Williams Jr., Joseph A.
/ REGISTRATION NUMBER: 38,659
/ REFERENCE/DOCKET NUMBER: 27866/32684
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312-474-6300
/ TELEFAX: 312-474-0448
/ TELEX: 25-3856
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1161 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-482-293A-2

Query Match      18.58; Score 1131.5; DB 2; Length 1161;
Best Local Similarity 29.48; Pred. No. 1e-99;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDEHHPLRPPGPEAEFGVSLQHVGGGORMWLVGAPWDPGSDRRGVDYRCPVGAH 82
DB 17 FNLVDEEPTIFQ-EDAGGFGSVVQF--GGR-LVVGAPLEVVAANTGLDYC---AAA 69
QY 83 NAPCAKGLHDYQLGNSSHPAVNMHLQMSLLETDGCGFMACAPLWBRACGSSVFSSGIC 142
DB 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASNTSGSRLLACGPTLHRVCGENSYSKGC 123
QY 143 ARVDASFOGSLAPTAQRCP-TYMDVVIVLDGNSI--YPMSEVQTFLLRLVQKLFIDP 199
DB 124 LLLGSRWEIITQTPDAPTECPHQEMDIVFLIDGSGSIDQNDFMQKGFVQAVMGQ--FEG 181
QY 200 EQIQVGLVQVGGESPVHWSLGDPRKEEVVRAAKNLSRREGRETAKTAQAIMVACTEGFSQ 259
DB 182 TDTLPAIMQVSNLLKIHFTTQFRTSPSQSLVDPIVQLKGL-TFTATGILTVVTQLFHH 240
QY 260 SHGRPEARLLVVVTDGESH-DGEELPAALKACEAGRVTTRYGIVAVLGHYLRQRDPSSF 318
DB 241 KNGARKSAKKILIVITDGQYKDPLEYSDDVIPQAEKAGIIRYAIGV-GHAF---QGPTA- 295
QY 319 LREIRTIASDPDRPFNFVTDAAALTDIVDALGDRIFGLEGSHANESSFGLMSQIGFS 378
DB 296 RQELNTISSAPPDQHVFKVDNFALGSIQKQLEKIYAVEGTQSRASSSFQHEMSQSGFS 355
QY 379 THRLLKQILFGMVGAYDWGSLWLEGGHRLFPPRMALEDEFPFALQNH-----AY 430
DB 356 TALTMDELFGVAGVSFN-----SGAFLYPPNMS-----PTFINMQENVDMRDSY 402
QY 431 LGYSVSMLLRGRRFLSGAPRHRGKVIATF-QLKKDGAVRVAQSLQGEIQISYFGSE 489
DB 403 LGYSTELALWKGVQNLVL-GAPRYQHTGKAVIFTQVSRQW--RKAQVETGTQIGSYFGAS 459
QY 490 LCPDTRDGTDTVDLLVAAPMFLGPQNKETGRVVYLV--GQQSLLTQGLTQPEPPQD- 546
DB 460 LCSVDVSDSGSDTDILIGAPHYY--EQTRGQGVSVCPFLPRGQVQWQCDVLRGEQGHWP 517
QY 547 ARFGFAGALPDNLQDGFADVAVGAPLEDGHQALYLVHG-TQSGVPHPAQRIATAASMP 605
DB 518 GRFQAALTVLGDVNNEDKLDIVAGAPGEQENRGAIVLFGHAGESGIGSFHSQRIASSQLS 577
QY 606 HALSYFGRSVDGRDLDDGDDLDVAVGAQGAAILLSRRPIVHLTPSLEVTPQAISVVQRD 665

578 PRLOVEGQALSGQDLTQDGLMDLAVGARGQVLLLRSLPVLKVGVMRSPVEVAKAVYR 637
666 C-----RRRGOEAVCLTAALCFQVTSRTPGRWHDHOFYMRFTASIDDEWTAGARAPDGS 718
638 CWEEKPSALEAGDATVCLTIQ-----KSLDQLGDIQSSVRFDLADLPGLTSRALFNET 692
719 GQRLSPRRRLRLSVGNVTCBQLHFPVLD-TSDYLRLPVALTVTTFALDNTTKPG-----PVLN 772
693 KNPTLTKRTKLG-LHCETLKLKLLPDCVEDVVSPIILHLNLSLVREPSPQNLRPVLA 751
773 EGSPTSIQKLVFPFKDQDNECVTDLVLOVNMIDIRGSRKAPVVRGRRKVLVSTTLEN 832
752 VGSQDLFTASLPFKNCGQDGLCEGD--LGVTLSPFSLQT---LTVGSSSLNLMNIVTVWN 806
833 RKENAYNTSLSIIFSRNLHLASLTPORESP-----IKVEC-AAPSAHARL-----CSVGHVP 883
807 AGEDSYGTVVSLYYPAGLSHRRVSGAQKQPHOSALRLACETVPTDEGLRSSKCSVNHP 866
884 FOTGAKVTFLLEPFEPFSCSLLSQVFGKLTASSDSLRNGTLOENTTAQTSAIYQYEPHLLP 943
867 FHGSGNGTPIVTFDVSFKATLGDRLM-LMRASASSENKASSKATQLELPVKYAVYVMI 925
944 SSESTLHRYEHPYGTLPVGPGEFKTTLRVQNLGYVVSGLIISALLPAVAHGMNYFLS 1003
926 SROESTKY--FNFATSDKMKKEAHRVYRVNLSQORDLA-ISINFVPVLLNGVAVMDV 982
1004 LSQVITNASCIVONLTPEPPVHPPEELOHNR---LNGSNTQCQVVRCHLQOLAKGTE 1060
983 VMEAPSQSIFPCVSR-----KPPQHSDFLTQISRSPMLDCSIADCLQFCDVPSFSVQEE 1037
1061 VSVGLLRLVHNEPFRRAKFKSLTVVSTFELGTGEGSVLQLTEASRWSESLLE-VVQTRPI 1119
1038 LDTLKGNSLPGWVRETLOKVLVSVVAEITPDTSVYSQLPGQEAFAWRAQEMVLESEDEV 1097
1120 LISLWLTIGSVLGLLALLLALLVCLMKLQFF-AHKKIPEEKREE 1163
1098 YNAIPITMGSSVGAULLLALITATLYKLGFFKFRHYKEMLEDKPED 1142

RESULT 14
US-08-943-363-2
; Sequence 2, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
```

```
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-943-363-2

Query Match      18.5%; Score 1131.5; DB 2; Length 1161;
Best Local Similarity 29.4%; Pred. No. 1e-99;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

Qy 23 FNLDEHPRLFPGPPEAEFGYSVLQHVGGQRMVLGAPWMDPGSDRRGDVYRCPVCGAH 82
Db 17 FNLQVEEPTIFQ-EDAGGFGQSVVQF--GGSRLVVGAPLEVAANQTGRLYDC---AAA 69
Qy 83 NAPCAKCHLGDYQLGNESHHPAVNMHLGMSLLETDGCGFMACAPLWGRACSSVFPSSGIC 142
Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASNGSRLLACGFTLHRVCGENSYSKGSC 123
Qy 143 ARVDASFPQGSAPTAQRCP-TYMDVVIVLDGNSI--YPSWEVQTFRLRLVCKLFDIP 199
Db 124 LLLGSRWEIQTVPDAPTECPHQEMDIVFLIDGSGSIDQDNFNQMGFVQVMQ--FEG 181
Qy 200 EQIQVLVQVGESPVHWSLGDFTKEEVVRAAKNLSRREGRETKTAQAIMVACTEGFSQ 259
Db 182 TDTLPAFMQSNLKIHFPTFTSPSQSLVDPPIVQLKGL-TFTATGILTVVTQLFHH 240
Qy 260 SHGRPEARLLVVTGESH-DEEELPAUKACEAGRTYTGIAVGHYLRORDPSSP 318
Db 241 KNGARKSAKKILVITDQKQKDPLEYSDDVIPQAEKAGIIRYAGV-CHAF---QGPTA- 295
Qy 319 LRETRTASDPDRFFNFVTEAALTDIVDALGRIFGLESHAENESSFGLEMSOIGFS 378
Db 296 RQELNTISSAPPQDHVFKVDNFALQSLOKQLEKIYAVEGTQGRASSFOHEMSQGF 355
Qy 379 THRLLKDGILFGWGDVGGSVLMLEGGHRLFFPRMALEDFPPALQNH-----AY 430
Db 356 TALTMQDGLFLGAVGFSW-----SGAFLYPPNMS-----PTFINMSQENVDMRDSY 402
Qy 431 LGYSVSMMLRGRRFLSGAPRRHRGKVIAF-OLKKDGAVRVAQSLQGEIQISYFGSE 489
Db 403 LGYSTELALMKGVQNLVL-GAPRYQHTGKAVIFTQVSRQW--RKKAQVETGTQIGSYFGAS 459
Qy 490 LCPDLDTRDGTVDLLVAAPMFLGPQNKETGRVVYLV--CQOSLLTLQGLTQPEPQD- 546
Db 460 LCSVDVDSGSDTDLILGAPHY--EOTRGQGVSVCPLRQGVQWQCDVLRGEQGHWP 517
Qy 547 ARFGFAMGALPDLNQDGFADVAVGAPLEDGQGALYLYHG-TQSGVRPHPAQRIAAASMP 605
Db 518 GRFGAALTVDGVNEDKLDVAIGAPGEQENRGAVILFHGASESGISPSHSQRIASSQLS 577
Qy 606 HALSYFGSRVDRLDLQGDLDLVAVAGAGQAAILLSRRPIVHLTPSLVETPQATSVQVORD 665
Db 578 PRLQYFGQALSGGQDLTDQGLMDLAVGARGQVLLRLSLPVLKGVAMRFSPEVAKAVYR 637
Qy 666 C-----RRQGEAVCLTAALCQVTSRTTGRWDHQFYMEFTASLDENTAGAAFDGS 718
Db 638 CWEEKPSALEAGDATVCLTIQ-----KSSLDQLDGIQSSVRFDLALDGLRITSRAIFNET 692
Qy 719 GORLSRRLRLSVGNVTCEQLHFVLD-TSDYLRPVALTVTFALDNTTKPG-----PVIN 772
Db 693 KNPTLTRKTLGLG-IHCETLKLILPDCVEDVSVFIIHLNFSILVREPIPSQNLRPVLA 751
```

```
773 EGSPTSIOKLVPRSKDCGPDNECVTDLVLOVNMIDIRSKAPFVVRGRRKVLVSTTLEN 832
Db 752 VGSODLFTASLPPEKNGCGDLCEGD--LGVTLFSFGLQT--LTVGSSLELNVIVTVWN 806
Qy 833 RKENAYNTSLSIIFSRNLHLASLTPQRESP-----IKVEC-AAPSAHARL-----CSVGHPV 883
Db 807 AGEDSYGTVVVLYYPAGLSHRRVSGAKQPHQSALRLACETVPTDEGLRSSRCSVNHP 866
Qy 884 FQTGAKVTFLLEPEFSCSSLLSQVFGKLTASSOSLERNGTLQENTQATSAIQIYEPHLLF 943
Db 867 FHEGNGTFTVTFDVSYKATLGDRLM-LMRASASENNKAASSKATFOLELPVKYAVTMI 925
Qy 944 SSESTLHRYEHPYGTLPVGPGEFTKTLRVQLNLCVYVSGSLIISALLPAVAHGNVFLS 1003
Db 926 SROEESTKY--FNPATSDKKMEAEHRYRVNLSQRDLA-ISINFWVPVLLINGVAVMDV 982
Qy 1004 LSQVITNNASCIQNLTEPPGPPVHPPELOHTNR---LNGSNTQCQVVRCHLGOLAKGTE 1060
Db 983 VMEAPSQSLFCVSEK-----KPPQHSDFLQISRPMLDCSIADCLQFRCDVPSPVQEE 1037
Qy 1061 VSVGLLRLVNEFPFRRAKFKSLTVVSTFELGTEGSGVLQLTEARWSESLLE--VVQTRPI 1119
Db 1038 LDFTLKNLSFGWVRETLOKKVLVSVVAEITFTDTSVLSQLPQGEAFMRAQMEMVLEDEV 1097
Qy 1120 LISLWILIGSVLGLLALLVFLCLWKLGRF-AHKKIPEEREKREE 1163
Db 1098 YNAIPIIMGSSVGALLLALLITATLYKLGFGRHYKEMLEDKPED 1142

RESULT 15
US-09-193-043-2
; Sequence 2, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 2
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-193-043-2

Query Match      18.5%; Score 1131.5; DB 3; Length 1161;
Best Local Similarity 29.4%; Pred. No. 1e-99;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

Qy 23 FNLDEHPRLFPGPPEAEFGYSVLQHVGGQRMVLGAPWMDPGSDRRGDVYRCPVCGAH 82
Db 17 FNLQVEEPTIFQ-EDAGGFGQSVVQF--GGSRLVVGAPLEVAANQTGRLYDC---AAA 69
Qy 83 NAPCAKCHLGDYQLGNSSHHPAVNMHLGMSLLETDGCGFMACAPLWGRACSSVFPSSGIC 142
Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASNGSRLLACGFTLHRVCGENSYSKGSC 123
Qy 143 ARVDASFPQGSAPTAQRCP-TYMDVVIVLDGNSI--YPSWEVQTFRLRLVCKLFDIP 199
Db 124 LLLGSRWEIQTVPDAPTECPHQEMDIVFLIDGSGSIDQDNFNQMGFVQVMQ--FEG 181
Qy 200 EQIQVLVQVGESPVHWSLGDFTKEEVVRAAKNLSRREGRETKTAQAIMVACTEGFSQ 259
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 6.89976 Seconds
(without alignments)
2761.097 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025

Sequence: 1 GICARVDASFPQGSGLAPTA.....FLREIRTIASDPDERFFNV 198

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	473	46.1	272	2 A55348	integrin alpha-1 -
2	467.5	45.6	1151	2 A45226	integrin alpha-1 c
3	465.5	45.4	1180	2 A35854	integrin alpha-1 c
4	454	44.3	1178	2 S44142	VLA-2 protein homo
5	451	44.0	1170	2 I45914	integrin alpha 2 s
6	441	43.0	1181	2 A33998	integrin alpha-2 c
7	336	32.8	191	2 I47230	VLA-2 protein - pi
8	227.5	22.2	1747	2 A45974	collagen alpha 1(X
9	227.5	22.2	1857	2 S31212	collagen alpha 1(X
10	227.5	22.2	1888	2 S78476	collagen alpha 1(X
11	227.5	22.2	3124	2 A40020	collagen alpha 1(X
12	214	20.9	493	2 A33809	cartilage matrix p
13	205	20.0	496	2 A37979	cartilage matrix p
14	204.5	20.0	929	2 I51027	type XII collagen
15	201.5	19.7	3137	2 A37797	hypothetical prote
16	199	19.4	3051	2 S42373	cartilage matrix p
17	194.5	19.0	500	2 S65522	cartilage matrix p
18	177.5	17.3	1179	2 A53213	integrin alpha-E c
19	176.5	17.2	3176	2 CGHU3A	collagen alpha 3(V
20	175	17.1	724	2 A48569	antigen Em100 - Ei
21	170.5	16.6	843	2 A40970	undulin 1 - human
22	169	16.5	741	2 T46488	hypothetical prote
23	158.5	15.5	1153	1 RWUUB	cell surface glyco
24	158	15.4	1163	2 I56126	lymphocyte fuction
25	157.5	15.4	1153	2 S00551	heukocyte surface
26	156	15.2	550	2 T23760	hypothetical prote
27	156	15.2	2944	2 A54849	collagen alpha 1(V
28	154	15.0	2813	1 VWUUC	von Willebrand fac
29	153.5	15.0	1163	1 RWUUC	cell surface glyco

30	152	14.8	1170	2 S03308	cell surface glyco
31	142.5	13.9	712	2 A45638	immunodominant mic
32	141	13.8	341	2 T32949	hypothetical prote
33	136.5	13.3	427	2 G00039	von Willebrand fac
34	129.5	12.6	414	2 PS0323	protein M01E10.2 (
35	123	12.0	1286	2 A88396	hypothetical prote
36	116.5	11.4	13055	2 T16580	hypothetical prote
37	114.5	11.2	567	2 T28797	hypothetical prote
38	112.5	11.0	661	2 T16597	hypothetical prote
39	110	10.7	643	2 T19549	hypothetical prote
40	107	10.4	449	2 T33257	protein CTRP - mal
41	103.5	10.1	2098	2 T18397	collagen alpha 2(V
42	100	9.8	1029	1 S21169	collagen alpha 1(V
43	97.5	9.5	780	2 A34102	collagen alpha 1(V
44	95.5	9.3	1019	1 A32856	undulin 2 - human
45	95	9.3	445	2 B40970	

ALIGNMENTS

RESULT 1

A55348

integrin alpha-1 - chicken (fragment)

C:Species: Gallus gallus (chicken)

C>Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004

C:Accession: A55348

R;Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.E.

J. Biol. Chem. 269, 22811-22816, 1994

A:Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1.

A:Reference number: A55348; MUID:94357930; PMID:7521332

A:Accession: A55348

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-272 <KER>

A:Cross-references: UNIPROT:O42094; GB:U10114

F;55-230/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 46.1%; Score 473; DB 2; Length 272;

Best Local Similarity 45.5%; Pred. No. 7,7e-36;

Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;

Qy	1	GICARVDASFPQGSGLAPTAQRCPYMDVIVLVDGNSIYPWSEVQTFRLRLVGLKFLDP	60
Db	30	GVCSNVSSTPETVKAVAPSVQCEKTDQDVIIVLDGNSIYPWSEVTAFLNSLLRNMDIGP	89
Qy	61	EQIQVGLVOYGESPVHEWSLGDPRTEKEVVRRAAKNLSRREGRETQTAQAINVACTEGFSQ	120
Db	90	QQTQVGIVQYQGVVHEFYLTNTYSTTEEVMDAALRIQRGGTQMTALGIDTAREEAFTE	149
Qy	121	SHGGRPEARLLVVVTVDGSHDGEELPAALKACAGRVTRYGVIAVLGHYLRQRDPSSFL	180
Db	150	AHGARRGVQVMVITVDGSHDNYRLQEDVDKCDENIQRFALILGYSRGNLSTKFFV	209
Qy	181	REIRTIASDPDERFFNV	198
Db	210	EIIKSIASKPTKHFNFV	227

RESULT 2

A45226

integrin alpha-1 chain - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: A45226

R;Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.

J. Biol. Chem. 269, 2989-2996, 1993

A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.

A:Reference number: A45226; MUID:93155124; PMID:8428973

A:Accession: A45226

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1151 <BRI>

A;Cross-references: UNIPROT:P56199
A;Experimental source: Hepatocarcinoma cell line HepG2
A;Note: sequence extracted from NCBI backbone (NCBIP:124326)
F;142-317/Domain: von Willebrand factor type A repeat homology <VWA1>

Query Match 45.6%; Score 467.5; DB 2; Length 1151;
Best Local Similarity 46.5%; Pred. No. 1.4e-34;
Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKLFIDP 60
Db 118 GICSDVPTFQVNSIAP-VQECSTQLDIVLDGNSIYPWSDVTAFLNLLKRMIDGP 176
Qy 61 EQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKQAQIMVACTEFSQ 120
Db 177 KQTQVIGVQYGENVTHFNLNKYSTBEVLVAANKIVQGRQRTWALGIDTARKEAFTE 236
Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 237 ARGARRGVKVMVITVDGSHDNRLKKVIQDCDENIQRFSAIILGSYNRGNLSTKFPV 296
Qy 181 REIRTIASDDPERFFNV 198
Db 297 EEIKSIASEPTKHFNV 314

RESULT 3
A35854
Integrin alpha-1 chain precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 23-Oct-1990 #sequence_revision 13-Sep-1991 #text_change 09-Jul-2004
C;Accession: A35854; S11243
R;Ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto,
J.; Cell Biol. 111, 709-720, 1990
A;Title: Molecular cloning of the rat integrin alpha-1 subunit: a receptor for laminin a
A;Reference number: A35854; MUID:90338125; PMID:2380249
A;Accession: A35854
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1180 <IGN>
A;Cross-references: UNIPROT:P18614; GB:X52140; NID:956493; PIDN:CAA36384.1; PID:G56494
C;Keywords: cell adhesion; cytoskeleton; transmembrane protein
F;170-345/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 45.4%; Score 465.5; DB 2; Length 1180;
Best Local Similarity 46.5%; Pred. No. 2.2e-34;
Matches 92; Conservative 35; Mismatches 70; Indels 1; Gaps 1;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKLFIDP 60
Db 146 GICSDVPTFQVNSIAP-VQECSTQLDIVLDGNSIYPWSEVIAFLNLLKRMIDGP 204
Qy 61 EQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKQAQIMVACTEFSQ 120
Db 205 KQTQVIGVQYGENVTHFNLNKYSTBEVLVAANKIQRGGLQRTWALGIDTARKEAFTE 264
Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 265 ARGARRGVKVMVITVDGSHDNRLKKVIQDCDENIQRFSAIILGSYNRGNLSTKFPV 324
Qy 181 REIRTIASDDPERFFNV 198
Db 325 EEIKSIASEPTKHFNV 342

RESULT 4
S44142
VLA-2 protein homolog - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S44142
R;Edelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latze
; submitted to the EMBL Data Library, January 1994

A;Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not
A;Reference number: S44142
A;Accession: S44142
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1178 <EDE>
A;Cross-references: UNIPROT:Q62469; EMBL:Z29987; NID:G473096; PIDN:CAA82877.1; PID:G47309
F;169-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 44.3%; Score 454; DB 2; Length 1178;
Best Local Similarity 46.0%; Pred. No. 2.5e-33;
Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKLFIDP 60
Db 144 GICSDVPTFQVNSIAP-VQECSTQLDIVLDGNSIYPWSEVIAFLNLLKRMIDGP 203
Qy 61 EQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKQAQIMVACTEFSQ 120
Db 204 KKTQVALIQYANEPRIIFNLNDPETKEDMVQATSETRQHGGLTNTFRAIEFARDVAYSQ 263
Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 264 TSGRPGATKVMVITVDGSHDGSGLKTVIQCCNDDEILRFGIAVLGYNLRNALDTKNLI 323
Qy 181 REIRTIASDDPERFFNV 198
Db 324 KEIKAIASPTERYFFNV 341

RESULT 5
I45914
Integrin alpha 2 subunit - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
C;Accession: I45914
R;Kamata, T.; Puzon, W.; Takada, Y.
J. Biol. Chem. 269, 9659-9663, 1994
A;Title: Identification of putative ligand binding sites within the I-domain of integrin
A;Reference number: A54402; MUID:94193647; PMID:7511592
A;Accession: I45914
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1170 <KAM>
A;Cross-references: UNIPROT:P53710; GB:L25886; NID:G439695; PIDN:AAB59255.1; PID:G439695
F;161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 44.0%; Score 451; DB 2; Length 1170;
Best Local Similarity 44.4%; Pred. No. 4.6e-33;
Matches 88; Conservative 37; Mismatches 73; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKLFIDP 60
Db 136 GICSDVPTFQVNSIAP-VQECSTQLDIVLDGNSIYPWSEVIAFLNLLKRMIDGP 195
Qy 61 EQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKQAQIMVACTEFSQ 120
Db 196 TKQMGILIQYANPRVFNLTFSKDEMIKATSTQFYQGGDLTNTFKAIQYARDTAYST 255
Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 256 AAGGRPGATKVMVITVDGSHDGSGLKKAVIDQCCKNILRFGIAVLGYNLRNALDTKNLI 315
Qy 181 REIRTIASDDPERFFNV 198
Db 316 KEIKAIASPTERYFFNV 333

RESULT 6
A33998
Integrin alpha-2 chain precursor - human
N;Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 ch
C;Species: Homo sapiens (man)

F/995-1076/Domain: fibronectin type III repeat homology <FN3G>	F/1086-1169/Domain: fibronectin type III repeat homology <FN3H>	F/1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>	F/1384-2295/Domain: IID #status predicted <IID>	F/1384-1465/Domain: fibronectin type III repeat homology <FN3I>	F/1474-1557/Domain: fibronectin type III repeat homology <FN3J>	F/1566-1647/Domain: fibronectin type III repeat homology <FN3K>	F/1655-1738/Domain: fibronectin type III repeat homology <FN3L>	F/1756-1838/Domain: fibronectin type III repeat homology <FN3M>	F/1847-1928/Domain: fibronectin type III repeat homology <FN3N>	F/1937-2019/Domain: fibronectin type III repeat homology <FN3O>	F/2028-2110/Domain: fibronectin type III repeat homology <FN3P>	F/2119-2199/Domain: fibronectin type III repeat homology <FN3Q>	F/2207-2294/Domain: fibronectin type III repeat homology <FN3R>	F/2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>	F/2438-2440/Region: cell adhesion #status predicted	F/2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted <COL2>	F/2751-2902/Domain: collagenous COL2 #status predicted <COL2>	F/2899-2901/Region: cell attachment (R-G-D) motif	F/2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>	F/2946-3048/Domain: collagenous COL1 #status predicted <COL1>	F/3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>	F/32, 1006.1032, 1044, 1512, 1767, 2210, 2273, 2532, 2683/Binding site: carboxyproline (Asn) (cova)	F/2780, 2789, 2836, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) (Hsta)	
Query Match	22.2%; Score 227.5; DB 2; Length 3124;	Best Local Similarity	34.1%; Pred. No. 4.9e-12;	Matches	59; Conservative	32; Mismatches	67; Indels	15; Gaps	5;															
Qy	28	DDVIVLDGNSI---YPMSEVOTFLRLVGLKFLDIPQIQVLVOYGESPVHEWSLGDFFRT	85	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Db	1199	DVILLVDGNSIGRPNFKFNIRISRIVEFDIGPKQVIGLAQYSGDPRTEWNLNAYRT	1258	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Qy	86	KEBVAANKLSRRGRETQTAAIIVACTEGFSQSHGGRPEARLLVVVTDGESHGDEE	145	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Db	1259	KEALLDRAVNLVYK-GENTUTGHALDPILKNVFKQAGLRPRARKIIVGLITDGKSQDDVV	1317	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Qy	146	LPAALKACEAGRVTRYGIAVLGHYLRQRDPSPFLREIRTIASDPDPPEFFNV	198	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Db	1318	TPSRLRDEG--VELVAIGI-----KNADN---ELKQIATDPDDIHAYNV	1358	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
RESULT 12																								
A33809																								
C:Species: Gallus gallus (chicken)																								
C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 09-Jul-2004																								
C:Accession: A33809; A26364																								
R:Kiss, I.; Deak, F.; Holloway Jr., R.G.; Delius, H.; Mebust, K.A.; Frimberger, E.; Argraves, W.S.; Deak, F.; Sparks, K.J.; Kiss, I.; Goetinck, P.F.																								
J. Biol. Chem. 264, 8126-8134, 1989																								
A:Title: Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix, von Willebrand factor, complement factors B and C2, and epidermal growth factor.																								
A:Reference number: A33809; MUID:89255246; PMID:254226																								

F:642-807/Domain: von Willebrand factor type A repeat homology <VM04>
F:840-1004/Domain: von Willebrand factor type A repeat homology <VM05>
F:1033-1197/Domain: von Willebrand factor type A repeat homology <VM06>
F:1237-1400/Domain: von Willebrand factor type A repeat homology <VM07>
F:1439-1604/Domain: von Willebrand factor type A repeat homology <VM08>
F:1639-1804/Domain: von Willebrand factor type A repeat homology <VM09>
F:1838-2010/Domain: von Willebrand factor type A repeat homology <VM10>
F:2043-2378/Domain: collagenous #status predicted <COL>
F:2045-2047/Region: cell attachment (R-G-D) motif
F:2153-2155/Region: cell attachment (R-G-D) motif
F:2159-2161/Region: cell attachment (R-G-D) motif
F:2379-3137/Domain: non-collagenous #status predicted <CNC>
F:2405-2577/Domain: von Willebrand factor type A repeat homology <VM11>
F:2623-2806/Domain: von Willebrand factor type A repeat homology <VM12>
F:2803-2846/Domain: platelet glycoprotein IB-like #status predicted <GPI>
F:2954-3039/Domain: fibronectin type III repeat homology <FN3>
F:3072-3122/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F:201,2084,2436,2563,2581,2683,2867,2920,3003/Binding site: carbohydrate (Asn) (covalent)

Query Match 19.7%; Score 201.5; DB 2; Length 3137;
Best Local Similarity 29.8%; Pred. No. 1.2e-09;
Matches 59; Conservative 42; Mismatches 72; Indels 25; Gaps 6;
Qy 2 ICARVDASFOQGS LAP-----TAQRCPTVMDVIVLDGNSI--YPSVQVFLRLV 53
Db 212 LVASVRTSMTPQAGAKLVKDI TAQES---ADLIFLDGSDNIGSVNFQAIRDFLVNLI 268
Qy 54 GKLFIDPEIQVGLVQYGESPVHEWSLGDPTKEEVVRAAKNLSRRREGRETAKAIVVA 113
Db 269 ESLRVGAQIHIHVGVYSDQPRTEFALNSYSTKADVLDAVKALSRFGKEANTGALEYV 328
Qy 114 CTGFSQSHGGRPEAA--RLIVVVDGSHDGEPLPAALKACEAGRTRYGIIVLGHVLR 171
Db 329 VENLFTQAGGSRIEBAVPQILVLISGGSSD--DIREGLAVKQASIFSISGVL----- 381
Qy 172 QRDPSSFLREIRTIASD 189
Db 382 -----NADSAELQIQIATD 394

Search completed: April 6, 2005, 12:17:03
Job time : 8.89977 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 29.3046 Seconds
(without alignments)
3459.921 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025

Sequence: 1 GICARVDASFOQGS LAPTA.....FLREIRTIASDPDERFFNV 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1025	100.0	1167	1 ITAG_HUMAN	O75578 homo sapien
2	547	53.4	1188	2 Q7QC3	Q7qc3 mus musculus
3	544	53.1	1189	1 ITAH_HUMAN	Q9ukx5 homo sapien
4	543	53.0	1188	1 ITAH_MOUSE	P61622 mus musculus
5	473	46.1	285	1 ITAI_CHICK	O90615 gallus gall
6	473	46.1	1171	2 O42094	O42094 gallus gall
7	467.5	45.6	1151	1 ITAI_HUMAN	P56199 homo sapien
8	465.5	45.4	1180	1 ITAI_RAT	P18614 rattus norv
9	454	44.3	1178	1 ITA2_MOUSE	Q62469 mus musculus
10	454	44.3	1178	2 Q6P1C7	Q6p1c7 mus musculus
11	451	44.0	1170	1 ITA2_BOVIN	P53710 bos taurus
12	441	43.0	1181	1 ITA2_HUMAN	P17301 homo sapien
13	389	38.0	323	2 Q8CB84	Q8cb84 mus musculus
14	336	32.8	191	2 Q29124	Q29124 sus scrofa
15	299	29.2	156	2 Q71V33	Q71v33 homo sapien
16	227.5	22.2	1888	1 CA1E_CHICK	P32018 gallus gall
17	227.5	22.2	3124	1 CA1C_CHICK	P13944 gallus gall
18	214.5	20.9	1797	2 Q80X19	Q80x19 mus musculus
19	214	20.9	493	1 CANA_CHICK	P05099 gallus gall
20	213.5	20.8	3119	1 CA1C_MOUSE	Q60847 mus musculus
21	209.5	20.4	3063	1 CA1C_HUMAN	Q99715 homo sapien
22	209	20.4	637	2 Q81VX1	Q81vx1 homo sapien
23	209	20.4	1284	2 Q6PI59	Q6pi59 homo sapien
24	207.5	20.2	1472	2 Q90ZAO	Q90zao gallus gall
25	207	20.2	821	2 Q6PYX2	Q6pyx2 brachydanio
26	205.5	20.0	1329	1 KF10_HUMAN	Q9p218 homo sapien
27	205	20.0	496	1 CANA_HUMAN	P21941 homo sapien
28	204.5	20.0	929	1 CA1C_NOTVI	Q91145 notophthalm
29	203.5	19.9	500	2 Q80VN5	Q80vn5 mus musculus
30	203.5	19.9	1816	2 Q8NT15	Q8nt15 brachydanio
31	201.5	19.7	3137	1 CA36_CHICK	P15989 gallus gall

32	201	19.6	490	2	O7SVT5	O7svt5 xenopus lae
33	199	19.4	3183	2	Q652C2	Q652c2 caenorhabdi
34	199	19.4	3767	1	MUA3_CABEL	P34576 caenorhabdi
35	197	19.2	810	2	Q8C435	Q8c435 mus musculus
36	196	19.1	956	1	MTN2_MOUSE	O08746 mus musculus
37	196	19.1	956	2	Q8R542	Q8r542 mus musculus
38	195.5	19.1	422	2	Q8K1I8	Q8k1i8 dipodomys m
39	194.5	19.0	500	1	CAMA_MOUSE	P51942 mus musculus
40	193.5	18.9	1160	2	Q8MKF4	Q8mkf4 felis silve
41	193	18.8	451	2	Q8N2G3	Q8n2g3 homo sapien
42	191	18.6	280	2	Q8C4R9	Q8c4r9 mus musculus
43	191	18.6	915	2	O6UWA5	O6uwa5 homo sapien
44	191	18.6	937	2	Q96FT5	Q96ft5 homo sapien
45	191	18.6	956	1	MTN2_HUMAN	O00339 homo sapien

ALIGNMENTS

RESULT 1

ID	ITAG_HUMAN	STANDARD;	PRT;	1167 AA.
AC	O75578; Q9UHZ8;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Integrin alpha-10 precursor.			
GN	Name=ITGA10;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Articular chondrocytes;			
RX	MEDLINE=98352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;			
RA	Camper L., Hellman U., Lundgren-Aakerlund E.,			
RT	"Isolation, cloning, and sequence analysis of the integrin subunit alpha10, a beta1-associated collagen binding integrin expressed on chondrocytes.";			
RL	J. Biol. Chem. 273:20383-20389(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Endothelial cells, and Heart;			
RX	MEDLINE=20169197; PubMed=10702680;			
RA	Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,			
RA	Wang S.-X., Langley R., Krissansen G.W.;			
RT	"The integrin alpha10 subunit: expression pattern, partial gene structure, and chromosomal localization.";			
RL	Cytogenet. Cell Genet. 87:238-244(1999).			
CC	-!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.			
CC	-!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10 associates with beta-1.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: Widely expressed with highest expression in muscle and heart. Found in articular cartilage.			
CC	-!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins with I-domains do not undergo protease cleavage.			
CC	-!- SIMILARITY: Belongs to the integrin alpha chain family.			
CC	-!- SIMILARITY: Contains 7 FG-GAP repeats.			
CC	-!- SIMILARITY: Contains 1 VWFA domain.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch>) or send an email to license@sib-sib.ch.

EMBL; AF074015; AAC31952.1; -
EMBL; AF112345; AAF21944.1; -
EMBL; AF172723; AAF61638.1; -

```
DR HSP; P18614; IMHP.
DR Genew; HGNC:6135; ITGA10.
DR MIM; 604042; -.
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0005518; P:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS02034; VWF; 1.
DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 1167 Integrin alpha-10.
FT DOMAIN 23 1122 Extracellular (Potential).
FT TRANSMEM 1123 1145 Potential.
FT DOMAIN 1146 1167 Cytoplasmic (Potential).
FT REPEAT 38 97 FG-GAP 1.
FT REPEAT 98 ? FG-GAP 2.
FT DOMAIN 167 350 VWF_A.
FT REPEAT 365 427 FG-GAP 3.
FT REPEAT 428 482 FG-GAP 4.
FT REPEAT 483 545 FG-GAP 5.
FT REPEAT 546 605 FG-GAP 6.
FT REPEAT 608 660 FG-GAP 7.
FT DOMAIN 1134 1140 Poly-Leu.
FT CA_BIND 494 502 Potential.
FT CA_BIND 558 566 Potential.
FT DISULFID 76 86 By similarity.
FT DISULFID 666 675 By similarity.
FT DISULFID 681 736 By similarity.
FT DISULFID 789 795 By similarity.
FT CARBOHYD 98 98 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 234 234 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 336 336 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 364 364 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 733 733 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 763 763 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 839 839 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 921 921 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 1011 1011 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 1018 1018 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 1039 1039 N-linked (GlcNAc... ) (Potential).
FT CONFLICT 844 844 I -> L (in Ref. 2).
FT CONFLICT 909 909 G -> V (in Ref. 2).
FT CONFLICT 926 926 E -> D (in Ref. 2).
SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25C1AE0 CRC64;

Query Match 100.0%; Score 1025; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 8e-82;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFPQGSGLAPTAQCPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 60
Db 140 GICARVDASFPQGSGLAPTAQCPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 199

Qy 61 EIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRTKTAQAINVACTEGFSQ 120
Db 200 EIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRTKTAQAINVACTEGFSQ 259

Qy 121 SHGGRPEARLLVVVTDGESHGDELPAALKACAGRVTRYGVIAVLGHLRQRDPSSFL 180
Db 260 SHGGRPEARLLVVVTDGESHGDELPAALKACAGRVTRYGVIAVLGHLRQRDPSSFL 319

Qy 181 REIRTIASDPDERFFNV 198
Db 181 REIRTIASDPDERFFNV 198

Query Match 100.0%; Score 547; DB 2; Length 1189;
Best Local Similarity 53.0%; Pred. No. 1.9e-39;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFPQGSGLAPTAQCPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 60
Db 137 GMCSSVNSNFRPSKTVAPALQRCQTYMDIVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 196

Qy 61 EIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRTKTAQAINVACTEGFSQ 120
Db 197 GQIQVGLVQYGEDAVDAHEFLNDYRSVKDVVEAAASHIEQRGQTETRTAFGEFARSEAFQ 256

Qy 121 SHGGRPEARLLVVVTDGESHGDELPAALKACAGRVTRYGVIAVLGHLRQRDPSSFL 180
Db 257 --GGRKCAKKVMIVITDGESHSDSPDLKVRQSEKQNVTRYAVAVLGYNNGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 3
ITAH HUMAN
ID ITAH HUMAN STANDARD; PRT; 1189 AA.
AC Q9UKX5; Q9UKQ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-11 precursor.
GN Name=ITGA11;
```



```

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 4
ID ITAH MOUSE STANDARD; PRT; 1188 AA.
AC P61622;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE Integrin alpha-11 precursor.
GN Name=ITGAL;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci F., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Brakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Kravinsky M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen (By
CC similarity).
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11
CC associates with beta-1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; BC058716; AAHS8716.1; -.
CC MGD; MGI:2442114; Itgal.
CC PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
CC PROSITE; PS02234; VWFA; 1.
CC Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
CC Repeat; Signal; Transmembrane.
CC SIGNAL 1 22 Potential.
CC CHAIN 23 1188 Integrin alpha-11.
CC DOMAIN 23 1141 Extracellular (Potential).
CC TRANSMEM 1142 1164 Potential.
CC DOMAIN 1165 1188 Cytoplasmic (Potential).
CC REPEAT 38 94 FG-GAP 1.
FT REPEAT 102 163 FG-GAP 2.
FT DOMAIN 164 345 VWFA.
FT REPEAT 359 420 FG-GAP 3.
FT REPEAT 422 475 FG-GAP 4.
FT REPEAT 477 537 FG-GAP 5.
FT REPEAT 539 598 FG-GAP 6.
FT REPEAT 601 653 FG-GAP 7.
FT CA_BIND 551 559 Potential.
FT CA_BIND 613 621 Potential.
FT DISULFID 76 83 By similarity.
FT DISULFID 121 139 Potential.
FT DISULFID 129 159 Potential.
FT DISULFID 659 668 By similarity.
FT DISULFID 674 729 By similarity.
FT DISULFID 781 787 By similarity.
FT DISULFID 881 893 By similarity.
FT CARBOHYD 82 82 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 95 95 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 291 291 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 331 331 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 358 358 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 449 449 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 462 462 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 528 528 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 642 642 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 694 694 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 857 857 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 894 894 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 973 973 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 1031 1031 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 1039 1039 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 1059 1059 N-linked (GLNAC. .) (Potential).
SQ SEQUENCE 1188 AA; 133011 MW; 080313C9D0D65422E CRC64;

Query Match 53.0%; Score 543; DB 1; Length 1188;
Best Local Similarity 52.5%; Pred. No. 4.2e-39;
Matches 104; Conservative 39; Mismatches 53; Indels 2; Gaps 1;

Qy 1 GICARVDASFOGSLAPTAQRCPTVMDVIVLDGNSIYPWSEVQTLRLVYKLFIDP 60
Db 137 GMCSTRANSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWSEVQHLNLIKPFYGP 196
Qy 61 EQQVGLVQVQGESPVHWSLGDPTKEEVVRAAKNLSRREGTKTAQAIMVACTEGFSQ 120
Db 197 GQLQVGIQVQGEDAVHEFLNDYRSVKDVVEASHIEQRCGTETRTAFGIEFARSEAFQK 256
Qy 121 SHGGRPEAARLLVVVTGDGSHDGEELPAALKACEAGRVTRYGIYLVHLYLRQRDPSSFL 180
Db 257 --GGRKGAKKVMIVITDGDGSHDSDPLEKVIQSEKDNVTRYAVAVLVGYNNRRGINPETFL 314
Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 5
ID ITAH_CHICK STANDARD; PRT; 285 AA.
AC Q90615;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a)
DE (Fragment).
GN Name=ITGAL;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.

```

RC TISSUE=Gizzard;
 RX MEDLINE=94357930; PubMed=7521332;
 RA Kern A., Briesewitz R., Bank I., Marcantonio E.B.;
 RT "The role of the I domain in ligand binding of the human integrin
 alpha 1 beta 1";
 RL J. Biol. Chem. 269:22811-22816(1994).
 CC -I- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and
 collagen. It recognizes the proline-hydroxylated sequence G-F-P-G-
 E-R in collagen.
 CC -I- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1
 associates with beta-1.
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -I- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
 with I-domains do not undergo protease cleavage.
 CC -I- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -I- SIMILARITY: Contains 1 VWFA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U10114; AAA59067.1; --
 DR HSP; P18614; IMHP.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; PARTIAL.
 DR PROSITE; PS0234; VWFA; 1.
 KW Cell adhesion; Glycoprotein; Integrin; Receptor; Transmembrane.
 FT NON_TER 1
 FT DOMAIN 1
 FT DOMAIN <1> 285 Extracellular (Potential).
 FT CARBOHYD 66 279 VWFA.
 FT CARBOHYD 2 2 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 40 40 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 208 208 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 232 232 N-linked (GlcNAc. . .) (Potential).
 FT NON_TER 285 285
 SQ SEQUENCE 285 AA; 31503 MW; 1B05D3246C5CA7E CRC64;
 Query Match 46.1%; Score 473; DB 1; Length 285;
 Best Local Similarity 45.5%; Pred. No. 1.2e-33;
 Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;
 Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 60
 Db 36 GVCNSVSTFETVKAVAPSVQECKTQLDIVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 95
 Qy 61 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRREGRETAKQAQIMVACTEGFSQ 120
 Db 96 QQTQVGIVQYQGVVHFEYLTNTYTTBEVMDAALRIQRGTQTMALGIDTAREEAFTE 155
 Qy 121 SHGCRPEAARLLVVVTGESHGDEELPAALACACAGRVTRYGIATVGLHYLRQRDPSSFL 180
 Db 156 AHGARRGVQKVMVITVDTGESHNDYRLQEVIDKCEDENIQRFATILGYSRGNLSTKFEV 215
 Qy 181 REINTIASDPPDERFFNV 198
 Db 216 EEIKSIASKTEKHFNNV 233
 RESULT 6
 O42094
 ID O42094 PRELIMINARY; PRT; 1171 AA.
 AC O42094;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Alpha integrin.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Gizzard;
 RX MEDLINE=97476270; PubMed=9334246; DOI=10.1074/jbc.272.42.26643;
 RA Obata H., Hayaashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,
 RA Sobue K.;
 RT "Smooth muscle cell phenotype-dependent transcriptional regulation of
 the alpha1 integrin gene";
 RL J. Biol. Chem. 272:26643-26651(1997).
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
 CC -I- SIMILARITY: Belongs to the integrin alpha chain family.
 DR EMBL; AB000470; BAA23160.1; --
 DR PIR; A55348; A55348.
 DR HSP; P18614; IMHP.
 DR GO; GO:0008305; C:integrin complex; IEA.
 DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
 DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 1.
 DR Pfam; PF00357; Integrin_alpha; 1.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 4.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS0234; VWFA; 1.
 KW Cell adhesion; Integrin; Transmembrane.
 SQ SEQUENCE 1171 AA; 130228 MW; B505A4D65F09736E CRC64;
 Query Match 46.1%; Score 473; DB 2; Length 1171;
 Best Local Similarity 45.5%; Pred. No. 6.7e-33;
 Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;
 Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 60
 Db 135 GVCNSVSTFETVKAVAPSVQECKTQLDIVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 194
 Qy 61 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRREGRETAKQAQIMVACTEGFSQ 120
 Db 195 QQTQVGIVQYQGVVHFEYLTNTYTTBEVMDAALRIQRGTQTMALGIDTAREEAFTE 254
 Qy 121 SHGCRPEAARLLVVVTGESHGDEELPAALACACAGRVTRYGIATVGLHYLRQRDPSSFL 180
 Db 255 AHGARRGVQKVMVITVDTGESHNDYRLQEVIDKCEDENIQRFATILGYSRGNLSTKFEV 314
 Qy 181 REINTIASDPPDERFFNV 198
 Db 315 EEIKSIASKTEKHFNNV 332
 RESULT 7
 ITAL_HUMAN
 ID ITAL_HUMAN STANDARD; PRT; 1151 AA.
 AC P56199;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
 GN Name=ITGAL;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=93155124; PubMed=8428973;
 RA Briesewitz R., Epstein M.R., Marcantonio E.E.;
 RT "Expression of native and truncated forms of the human integrin alpha
 1 subunit.";
 RL J. Biol. Chem. 268:2989-2996(1993).
 CC -!- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and
 CC collagen. It recognizes the proline-hydroxylated sequence G-P-P-G-
 CC E-R in collagen.
 CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1
 CC associates with beta-1.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DOMAIN: The integrin I-domain (insert) is a VMPA domain. Integrins
 CC with I-domains do not undergo protease cleavage.
 CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -!- SIMILARITY: Contains 1 VMPA domain.
 CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD49a entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49a.htm".
 DR PTR; A45226; A45226.
 DR PDB; 1QC5; X-ray; A/B=140-331.
 DR PDB; 1QCY; X-ray; A=141-333.
 DR Genew; HGNC:6134; ITGAL.
 DR MIM; 192868; .
 DR GO; GO:0008305; C:integrin complex; TAS.
 DR GO; GO:0005518; F:collagen binding; TAS.
 DR GO; GO:0007160; P:cell-matrix adhesion; NAS.
 DR InterPro; IPR004113; Integrin_alpha.
 DR InterPro; IPR02035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 3.
 DR Pfam; PF00357; Integrin_alpha; 1.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR01185; INTEGRIN.
 DR PRINTS; PR00453; VMPADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA_1.
 DR PROSITE; PS00242; INTEGRIN ALPHA; 1.
 DR PROSITE; PS0234; VMPA; 1.
 DR 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;
 KW Magnesium; Receptor; Repeat; Transmembrane.
 FT DOMAIN 1 1113 Extracellular (Potential).
 FT TRANSMEM 1114 1136 Potential.
 FT DOMAIN 1137 1151 Cytoplasmic (Potential).
 FT REPEAT 16 75 FG-GAP 1.
 FT REPEAT 76 ? FG-GAP 2.
 FT REPEAT 147 360 VMPA.
 FT REPEAT 349 404 FG-GAP 3.
 FT REPEAT 405 457 FG-GAP 4.
 FT REPEAT 459 520 FG-GAP 5.
 FT REPEAT 540 599 FG-GAP 6.
 FT REPEAT 602 654 FG-GAP 7.
 FT CA_BIND 470 478 Potential.
 FT CA_BIND 552 560 Potential.
 FT CA_BIND 614 622 Potential.
 FT SITE 1139 1142 GFFKR motif.
 FT DISULFID 54 64 By similarity.
 FT DISULFID 660 669 By similarity.
 FT DISULFID 675 728 By similarity.
 FT DISULFID 780 786 By similarity.
 FT DISULFID 850 858 By similarity.
 FT DISULFID 1002 1034 By similarity.
 FT DISULFID 1037 1044 By similarity.
 FT CARBOHYD 46 46 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 72 72 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 77 77 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 84 84 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 189 189 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 289 289 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 313 313 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 374 374 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 390 390 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 432 432 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 504 504 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 671 671 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 720 720 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 752 752 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 812 812 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 855 855 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 880 880 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 887 887 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 911 911 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 938 938 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 946 946 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 980 980 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 1045 1045 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 1055 1055 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 1074 1074 N-linked (GlcNAc. .) (Potential).
 FT CARBOHYD 1085 1085 N-linked (GlcNAc. .) (Potential).
 FT STRAND 143 150
 FT TURN 153 154
 FT TURN 158 158
 FT HELIX 159 169
 FT TURN 170 171
 FT STRAND 174 174
 FT TURN 176 177
 FT STRAND 179 186
 FT STRAND 190 194
 FT TURN 196 197
 FT HELIX 202 211
 FT HELIX 222 231
 FT TURN 232 234
 FT HELIX 236 238
 FT TURN 239 239
 FT TURN 242 243
 FT STRAND 245 252
 FT TURN 257 258
 FT HELIX 259 261
 FT HELIX 262 271
 FT TURN 272 273
 FT STRAND 274 281
 FT HELIX 283 287
 FT TURN 288 289
 FT HELIX 293 302
 FT HELIX 307 310
 FT STRAND 311 314
 FT HELIX 317 323
 FT HELIX 324 330
 SQ SEQUENCE 1151 AA; 127837 MW; 6B3F3C1AABF52808 CRC64;
 Query Match 45.6%; Score 467.5; DB 1; Length 1151;
 Best Local Similarity 46.5%; Pred. No. 2e-32;
 Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;
 Qy 1 GICARVDASFPQGS LAPTAQRCPYMDVVIVLDGSGNSIYPWSEVQTFLLRLVGLKLFIDP 60
 Db 118 GICSDVSPTFQVNSIAP-VQECSTQLDIVLDGSGNSIYPWSDVTAFLNDLLKRWIDGP 176
 Qy 61 EQIQVGLVQYGESPVHWSLGDFTKBEVVVRAAKNLSRRREGTKTAQAIWVACTEFSQ 120
 Db 177 KQTQGVIVQYGENVTHFNLKYSSTEEVLVAAKKIVQGRGQTMTALGDTTARKEAFTE 236
 Qy 121 SHGGRPEAARLLVVVTDGESHGDBELPAALKACEAGRVTVYGIATLVGHYLRQRDPSSFL 180
 Db 237 ARGARGVKVMVITVDGESHNHLKVKVQDCEDENIQRFSTAILGSGYNRGNLSKTFV 296
 Qy 181 REIRTIASDPDERFFNV 198
 Db 297 EEIKSIASEPTEKHFFNV 314
 RESULT 8
 ID ITAI_RAT STANDARD; PRT; 1180 AA.
 AC P18614;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)

Integrin alpha-1 precursor (Laminin and collagen receptor) (VLA-1) (CD49a).

Name=Itgal;

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

NCBI_TaxID=10116;

[1]

SEQUENCE FROM N.A.

RP MEDLINE=90338125; PubMed=2380249; DOI=10.1083/jcb.111.2.709;

RX Ignatovic M.J., Large T.H., Houde M., Tawil J.W., Barton A., Esch F.,

RA Carbonetto S., Reichardt L.F.;

RA "Molecular cloning of the rat integrin alpha 1-subunit: a receptor for

RT laminin and collagen.";

RL J. Cell Biol. 111:709-720(1990).

[2]

RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.

RP MEDLINE=9313197; PubMed=10386626; DOI=10.1016/S0014-5793(99)00666-3;

RX Nolte M., Pepinsky R.B., Venyaminov S.Y., Kotliansky V.,

RA Gotwals P.J., Karpusas M.;

RA "Crystal structure of the alphabeta1 integrin I-domain: insights into

RT integrin I-domain function.";

RL FEBS Lett. 452:379-385(1999).

CC -!- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and

CC collagen. It recognizes the proline-hydroxylated sequence G-F-P-G-

CC E-R in collagen.

CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1

CC associates with beta-1.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins

CC with I-domains do not undergo protease cleavage.

CC -!- SIMILARITY: Belongs to the integrin alpha chain family.

CC -!- SIMILARITY: Contains 7 FG-GAP repeats.

CC -!- SIMILARITY: Contains 1 VWFA domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X52140; CAA36384.1; -.

DR PIR; A35854; A35854.

DR PDB; 1CK4; X-ray; A/B=167-364.

DR PDB; 1MHP; X-ray; A/B=169-360.

DR RGD; 2923; Itgal.

DR InterPro; IPR000413; Integrin_alpha.

DR InterPro; IPR002035; VWFA.

DR Pfam; PF01839; FG-GAP; 3.

DR Pfam; PF00357; Integrin_alpha; 1.

DR Pfam; PF00092; VWFA; 1.

DR PRINTS; PR01185; INTEGRINA.

DR PRINTS; PR00453; VWFADOMAIN.

DR SMART; SM00191; Int_alpha; 5.

DR SMART; SM00327; VWFA; 1.

DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.

DR PROSITE; PS50234; VWFA; 1.

DR 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;

KW Magnesium; Receptor; Repeat; Signal; Transmembrane.

FT SIGNAL 1 28

FT CHAIN 29 1180 Integrin alpha-1.

FT DOMAIN 29 1142 Extracellular (Potential).

FT TRANSMEM 1143 1165 Potential.

FT DOMAIN 1166 1180 Cytoplasmic (Potential).

FT REPEAT 44 103 FG-GAP 1.

FT REPEAT 104 ? FG-GAP 2.

FT DOMAIN 175 388 VWFA.

FT REPEAT 377 432 FG-GAP 3.

FT REPEAT 433 484 FG-GAP 4.

FT REPEAT 485 565 FG-GAP 5.

FT REPEAT 567 626 FG-GAP 6.

FT REPEAT 629 681

FT CA_BIND 497 505 Potential.

FT CA_BIND 579 587 Potential.

FT CA_BIND 641 649 Potential.

FT SITE 1168 1172 GPFKR motif.

FT DISULFID 82 92 By similarity.

FT DISULFID 687 696 By similarity.

FT DISULFID 702 755 By similarity.

FT DISULFID 807 813 By similarity.

FT DISULFID 877 885 By similarity.

FT DISULFID 1029 1062 By similarity.

FT DISULFID 1066 1073 By similarity.

FT CARBOHYD 100 100 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 105 105 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 112 112 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 217 217 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 317 317 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 341 341 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 402 402 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 418 418 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 459 459 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 531 531 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 698 698 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 747 747 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 779 779 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 820 820 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 839 839 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 882 882 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 907 907 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 938 938 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 965 965 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 973 973 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 1007 1007 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 1084 1084 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 1103 1103 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 1114 1114 N-linked (GlcNAc. .) (Potential).

FT STRAND 172 178

FT HELIX 186 197

FT TURN 198 199

FT TURN 204 205

FT STRAND 208 214

FT STRAND 218 222

FT TURN 224 226

FT HELIX 230 238

FT TURN 239 239

FT HELIX 250 259

FT TURN 260 261

FT HELIX 264 266

FT TURN 267 267

FT TURN 270 271

FT STRAND 274 280

FT TURN 285 286

FT HELIX 287 289

FT HELIX 290 299

FT TURN 300 301

FT STRAND 302 309

FT HELIX 311 315

FT TURN 316 317

FT HELIX 321 328

FT TURN 329 330

FT HELIX 335 338

FT STRAND 339 342

FT HELIX 346 350

FT HELIX 352 359

SQ SEQUENCE 1180 AA; 130808 MW; 8E5DA2BE02362EE4 CRC64;

Query Match 45.4%; Score 465.5; DB 1; Length 1180;

Best local Similarity 46.5%; Pred. No. 3.le-32;

Matches 92; Conservative 35; Mismatches 70; Indels 1; Gaps 1;

Qy 1 GICARDASFPQGSGLAPTRCPTVMDVIVLDGNSIYPWSEVQVFLRELCKLFDIP 60

Db 146 GICSDVSPTFQVNSFAP-VQECSTQLDVIIVLDGNSIYPWSEVIAFLNLLKRMIDIGP 204

```

Qy 61 EQIQVGLVQGESPVHWSLGDFTKBEVRAAKNLSRRREGRETAKTAQAIWVACTGFSQ 120
Db 205 KQTQGVQVQGVNTHEFNLYKSTSEVLVAANKIGRQGLQMTALGIDTARKEAFTE 264
Qy 121 SHGGRPBAARLLVVVTGESHGDBELPAALAKACBAGRVTRYGIAVLGHYLRQRDPSFL 180
Db 265 ARGARRGVKKVMVVTGESHGDNRLKQVQDCEDENIQRFSTAILGHYNRGNLSTKFPV 324
Qy 181 REIRTIASDPDERFFNV 198
Db 325 EEIKSIASEPTEKHFNV 342

RESULT 9
ITA2 MOUSE STANDARD; PRT; 1178 AA.
AC Q62469; 062163;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
GN Name=Itga2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=94363406; PubMed=8081889;
RA Edelman J.M., Chan B.M., Unival S., Onodera H., Wang D.Z.,
RA Danjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;
RT "The mouse VLA-2 homologue supports collagen and laminin adhesion but
RL not virus binding.";
RN Cell Adhes. Commun. 2:131-143(1994).
[2]
RP SEQUENCE OF 450-1178 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=94355691; PubMed=7521231;
RA Wu J.B., Santoro S.A.;
RT "Complex patterns of expression suggest extensive roles for the alpha
RL 2 beta 1 integrin in murine development.";
Dev. Dyn. 199:292-314(1994).
CC 1- FUNCTION: Integrin alpha-2/beta-1 is a collagen receptor, being
CC responsible for adhesion of platelets and other cells to
CC collagen, modulation of collagen and collagenase gene expression,
CC force generation and organization of newly synthesized
CC extracellular matrix. It is also a receptor for laminins, collagen
CC C-propeptides and E-cadherin. Mice homozygous for a null mutation
CC in the alpha-2 die very early in embryogenesis.
CC 1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
CC associates with beta-1.
CC 1- SUBCELLULAR LOCATION: Type I membrane protein.
CC 1- DOMAIN: The integrin I-domain (insert) is a VMPA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC 1- SIMILARITY: Belongs to the integrin alpha chain family.
CC 1- SIMILARITY: Contains 7 FG-GAP repeats.
CC 1- SIMILARITY: Contains 1 VMPA domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z29987; CAA82877.1; -
EMBL; X75427; CAA53178.1; -
PIR; S44142; S44142.
HSSP; P17301; IAOX.

```

```

DR MGD; MGI:96600; Itga2.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP_2.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWF_A; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;
KW Receptor; Repeat; Signal; Transmembrane.
FT SIGNAL 1 26 By similarity.
FT CHAIN 27 1178 Integrin alpha-2.
FT DOMAIN 27 1129 Extracellular (Potential).
FT TRANSMEM 1130 1151 Potential.
FT DOMAIN 1152 1178 Cytoplasmic (Potential).
FT REPEAT 42 100 FG-GAP 1.
FT REPEAT 101 7 FG-GAP 2.
FT DOMAIN 185 375 VMPA.
FT REPEAT 376 430 FG-GAP 3.
FT REPEAT 431 483 FG-GAP 4.
FT REPEAT 485 546 FG-GAP 5.
FT REPEAT 548 607 FG-GAP 6.
FT REPEAT 612 664 FG-GAP 7.
FT CA_BIND 496 504 Potential.
FT CA_BIND 560 568 Potential.
FT CA_BIND 624 632 Potential.
FT SITE 480 482 Cell attachment site (Potential).
FT SITE 1154 1158 GPPK motif.
FT DISULFID 80 89 By similarity.
FT DISULFID 677 734 By similarity.
FT DISULFID 786 792 By similarity.
FT DISULFID 862 873 By similarity.
FT DISULFID 1016 1047 By similarity.
FT DISULFID 1052 1057 By similarity.
FT CARBOHYD 102 102 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 109 109 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 429 429 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 457 457 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 472 472 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 696 696 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1054 1054 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1071 1071 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1078 1078 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;

Query Match 44.3%; Score 454; DB 1; Length 1178;
Best Local Similarity 46.0%; Pred. No. 3-2e-31;
Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;

Qy 1 GICARVDASPOQGSAPTACRQCTYMDVVVILDGNSIYPMSEVQTFRLRLVGLKFLDP 60
Db 144 GICSDVSPDFQFLTSFPAQVACPSLVDVVVCDSENSIYPMSEAVKNFLVKFTGLDIGP 203
Qy 61 EQIQVGLVQGESPVHWSLGDFTKBEVRAAKNLSRRREGRETAKTAQAIWVACTGFSQ 120
Db 204 KKTQVALIQVANEPRIFNLNDFETKEDMVOQTSFTRQHGDDLNTNTRAFEFARDVAYSQ 263
Qy 121 SHGGRPBAARLLVVVTGESHGDBELPAALAKACBAGRVTRYGIAVLGHYLRQRDPSFL 180
Db 264 TSGRPGATKVMVVTGESHGDSKLTQVQQNDDEILAFGLVGLYNALDTQNL 323
Qy 181 REIRTIASDPDERFFNV 198
Db 324 KEIKAIASPTETRYFFNV 341

RESULT 10
Q6PIC7 PRELIMINARY; PRT; 1178 AA.
ID Q6PIC7

```

```
AC 06P1C7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Integrin alpha 2.
GN Name=Itga2;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Eye;
RX MEDLINE=223825; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Goughman J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Eye;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; BC065113; AAH65139.1; --
DR HSSP; P17301; IAOX.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 2_
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA_1
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
DR Cell adhesion; Integrin; Transmembrane.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 11178 AA; 128954 MW; 62FAEA820242A9B6 CRC64;

Query Match 44.3%; Score 454; DB 2; Length 1178;
Best Local Similarity 46.0%; Pred. No. 3.2e-31;
Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGSAPTAQRCPTMDVVLVLDGNSIYPWSEVQTFELRLVGLKFLDIP 60
Dy 144 GICSDVDFQFLTSFSPAQACSLVDVVVCDENSIYPWEAVKFLVKFVGLDIP 203
Qy 61 EQIQVLVQGESPVHESGLDFTKEEVRAAKNLSRREGRTKTAQIMVACTGFSQ 120
Dy 204 KKTQVALIQYANEPRIIFNLNDFETKEDMWQATSETRQHGGLDITNFAIEFARDYAYSQ 263
Qy 121 SHGGRPEARLLVVVTDGSHGSLKLVITQCCNDDEILRFGIAGVLGNRLDITKMLI 323
```

```
Db 264 TSGRPGATKVMVVVTDGSHGSLKLVITQCCNDDEILRFGIAGVLGNRLDITKMLI 323
Qy 181 REINTIASDPDERFFNV 198
Dy 324 KEIKAIASPTPTERYFFNV 341

RESULT 11
ID ITA2_BOVIN STANDARD; PRT; 1170 AA.
AC P53710;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).
GN Name=ITGA2;
OS Bos taurus (Bovine);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94193647; PubMed=7511592;
RA Kamata T., Puzon W., Takada Y.;
RT "Identification of putative ligand binding sites within I domain of
RL integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";
RL J. Biol. Chem. 269:9659-9663(1994).
CC -1- FUNCTION: Integrin alpha-2/beta-1 is a receptor for laminin,
CC collagen, collagen C-propeptides, fibronectin and E-cadherin. It
CC recognizes the proline-hydroxylated sequence G-P-G-E-R in
CC collagen. It is responsible for adhesion of platelets and other
CC cells to collagens, modulation of collagen and collagenase gene
CC expression, force generation and organization of newly synthesized
CC extracellular matrix.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
CC associates with beta-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: The integrin I-domain (insert) is a VWF domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VWF domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; L25886; AAB59255.1; --
DR PIR; I45914; I45914.
DR HSSP; P17301; IAOX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 2_
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA_1
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;
KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.
FT NON_TER 1 1
FT SIGNAL <1 18 Potential.
FT CHAIN 19 1170 Integrin alpha-2.
FT DOMAIN 19 1121 Extracellular (Potential).
```


DR PDB; 1DZ1; X-ray; A=171-355.
 DR Genew; HGNC:6137; ITGA2.
 DR MM; 192374; .
 DR GO; GO:0008305; C:integrin complex; TAS.
 DR GO; GO:0005886; C:plasma membrane; TAS.
 DR GO; GO:0005518; F:collagen binding; TAS.
 DR GO; GO:0007596; P:blood coagulation; TAS.
 DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
 DR GO; GO:0009887; P:organogenesis; TAS.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 2.
 DR Pfam; PF00357; Integrin_alpha; 1.
 DR PRINTS; PRO1185; INTEGRIN.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS0234; VWFA; 1.
 DR 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;
 KW Glycoprotein; Integrin; Magnesium; Platelet; Polymorphism; Receptor;
 KW Repeat; Signal; Transmembrane.
 FT SIGNAL 1 29
 FT CHAIN 30 1181 Integrin alpha-2.
 FT DOMAIN 1133 1154 Extracellular (Potential).
 FT TRANSMEM 1155 1181 Potential.
 FT DOMAIN 1155 1161 Cytoplasmic (Potential).
 FT REPEAT 45 103 Interaction with HP85.
 FT REPEAT 104 7 FG-GAP 1.
 FT REPEAT 104 7 FG-GAP 2.
 FT DOMAIN 188 378 VWFA.
 FT REPEAT 379 433 FG-GAP 3.
 FT REPEAT 434 486 FG-GAP 4.
 FT REPEAT 488 549 FG-GAP 5.
 FT REPEAT 551 610 FG-GAP 6.
 FT REPEAT 615 667 FG-GAP 7.
 FT CA_BIND 499 507 Potential.
 FT CA_BIND 563 571 Potential.
 FT CA_BIND 627 635 Potential.
 FT SITE 1157 1161 GFPR motif.
 FT DISULFID 83 92 By similarity.
 FT DISULFID 680 737 By similarity.
 FT DISULFID 789 795 By similarity.
 FT DISULFID 865 876 By similarity.
 FT DISULFID 1019 1050 By similarity.
 FT DISULFID 1055 1060 By similarity.
 FT CARBOHYD 105 105 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 112 112 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 343 343 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 432 432 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 460 460 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 475 475 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 699 699 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1057 1057 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1074 1074 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1081 1081 N-linked (GlcNAc...) (Potential).
 FT VARIANT 534 534 K -> E (in alloantigen HPA-5B;
 dbSNP:1801106).
 FT TURN 170 171 /FTid=VAR_003977.
 FT STRAND 173 180
 FT TURN 183 184
 FT HELIX 188 199
 FT TURN 200 201
 FT STRAND 204 204
 FT TURN 206 207
 FT STRAND 209 216
 FT STRAND 220 224
 FT TURN 226 228
 FT HELIX 232 240
 FT TURN 241 241
 FT HELIX 252 262

FT TURN 263 264
 FT HELIX 266 268
 FT TURN 269 269
 FT STRAND 275 282
 FT HELIX 289 291
 FT HELIX 292 301
 FT TURN 302 303
 FT STRAND 304 311
 FT HELIX 313 317
 FT TURN 318 319
 FT HELIX 323 330
 FT TURN 331 332
 FT HELIX 337 340
 FT STRAND 341 344
 FT HELIX 347 353
 FT HELIX 354 362
 FT TURN 363 363
 SQ SEQUENCE 1181 AA; 129295 MW; 7E1B7ED968A94070 CRC64;
 Query Match 43.0%; Score 441; DB 1; Length 1181;
 Best local Similarity 44.4%; Pred. No. 4.6e-30;
 Matches 88; Conservative 33; Mismatches 77; Indels 0; Gaps 0;
 QY 1 GICARVDASFOQGSILAPTAQRCPYMDVIVLDGNSIYPWSEVQTFRLRLVCKLFDIP 60
 DB 147 GVCSDSIPDFQLSASFSPATQPCPSLIDVVVCDENSIYPWDVAVKVFLEKVGQLDIGP 206
 QY 61 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRREGRETAKTAQAINVACTEGFSQ 120
 DB 207 TKQVGLIQYANNPRVVFNLTNTYKTEEMIVATVTSQSYGGDLTNTFGAIQYARKAYSA 266
 QY 121 SHGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIIVGLHYLRQRDPSSFL 180
 DB 267 ASGGRSATKVMVVVTDGSHDGSMLKAVIDQCNDHNLIRPGIAVLGYLNNRLDTRKLI 326
 QY 181 REIRTIASDPDERFFNV 198
 DB 327 KEIKAIASIPTRYFFNV 344
 RESULT 13
 ID Q8CB84 PRELIMINARY; PRT; 323 AA.
 AC Q8CB84; 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Mus musculus adult male bone cDNA, RIKEN full-length enriched library,
 clone:9830137A04 product:integrin alpha 2, full insert sequence.
 DE (Fragment).
 DE Name=Itga2;
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Bone;
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Bone;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Bone;
 RA The FANTOM Consortium,

DR Pfam: PF00092; VWA; 1.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; VWA; 1.
KW Integrin.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 17042 MW; BC7ABD63226C652B CRC64;

Query Match 29.2%; Score 299; DB 2; Length 156;
Best Local Similarity 44.1%; Pred. No. 1.5e-18;
Matches 63; Conservative 24; Mismatches 56; Indels 0; Gaps 0;

Qy	56	LFDIDPEIQVGLVQYGESPVHWSLGDFTKKEEVVRAAKNLSRREGRETKTQAQIMVACT	115
Db	3	LDIGPTKTQVGLIQYANNPRVFNLTNTYKTEEMIVATSTQSYGGDLTNTFGAIQYARK	62
Qy	116	EGFSQSHGGRPEAARLLVWVDGSHDGEELPAALKACEAGRVTTRYGIIVLGHVLRQRD	175
Db	63	YAYSASGGRRSATKVMVVVDGSHDGSMLKKAVIDQCNHDNILRFGIIVLGYLNRNALD	122
Qy	176	PSSFLREIRTIASDPDERFFNV	198
Db	123	TKNLIKSIAISIPTRYFFNV	145

Search completed: April 6, 2005, 12:15:27
Job time : 31.3046 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 25.0407 Seconds
(without alignments)
3058.161 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025

Sequence: 1 GICARVDASFOGSLAPTA.....FLREIRTIASDPDERFFNV 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1025	100.0	1132	3 AAY32243	Aay32243 Human int
2	1025	100.0	1152	4 AAB64657	Aab64657 Human sec
3	1025	100.0	1152	4 AAB64658	Aab64658 Human sec
4	1025	100.0	1167	3 AAY32242	Aay32242 Human int
5	1025	100.0	1167	4 AAB64584	Aab64584 Human sec
6	1025	100.0	1167	6 ABP99490	Abp99490 Human sec
7	1025	100.0	1167	6 ABR00964	Abr00964 Human gen
8	1025	100.0	1167	6 ADA44026	Ada44026 Human sec
9	1025	100.0	1167	8 ADQ19290	Adq19290 Human sof
10	912	89.0	195	5 AAU76862	Aau76862 Human int
11	912	89.0	195	5 AAU76853	Aau76853 Human int
12	912	89.0	198	8 ADR23227	Adr23227 Human int
13	840.5	82.0	1049	5 ADR41424	Adr41424 Human CD-
14	797.5	77.8	303	3 AAY32282	Aay32282 Mouse int
15	547	53.4	1188	4 AAB50087	Aab50087 Murine A2
16	547	53.4	1188	5 AAU10552	Aau10552 Murine A2
17	544	53.1	707	4 AAU19663	Aau19663 Human nov
18	544	53.1	707	5 ABP47883	Abp47883 Human pol
19	544	53.1	707	7 ADC10845	Adc10845 Human ext
20	544	53.1	1034	3 AAB25590	Aab25590 Protein e
21	544	53.1	1034	6 ADA27062	Ada27062 Human nov
22	544	53.1	1034	8 ADE86592	Ade86592 Novel hum
23	544	53.1	1058	5 ADR41496	Adr41496 Human CD-
24	544	53.1	1188	4 AAB30929	Aab30929 Amino aci
25	544	53.1	1188	4 AAU14467	Aau14467 Human nov

ALIGNMENTS

RESULT 1

AAAY32243

ID AAY32243 standard; protein; 1132 AA.

XX AC AAY32243;

XX DT 15-FEB-2000 (first entry)

XX DE Human integrin subunit alpha-10 splice variant.

XX KW Integrin alpha-10; ISal0; human; trauma; rheumatoid arthritis;

XX KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;

XX KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker;

XX KW splice variant.

XX OS Homo sapiens.

XX PH Key

FT Peptide

FT Protein

FT /note= "signal peptide"

FT /note= "mature protein"

XX W09951639-A1.

XX 14-OCT-1999.

XX 31-MAR-1999; 99WO-SE0000544.

XX 02-APR-1998; 98SE-00001164.

XX 28-JAN-1999; 99SE-00000319.

XX (ACTI-) ACTIVE BIOTECH AB.

XX Lundgren-Akerlund E;

XX WPI; 2000-052639/04.

XX N-PSDB; AAZ34720.

XX New isolated integrin subunit alpha-10, used as a marker or target

XX molecule for cells during development, regeneration and pathological

XX conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or

XX inflammation.

XX Claim 1; Page 43-48; 90pp; English.

XX This sequence represents a splice variant of novel human chondrocyte

XX integrin subunit alpha-10 (ISal0). It is identical to ISal0 (see

CC AAY32242) except for deletion of amino acids 975-986. The invention
CC relates to a recombinant or isolated integrin heterodimer comprising the
CC alpha10 subunit in association with subunit beta (especially beta-1). The
CC heterodimer, subunit alpha-10 or splice variant can be used as a marker
CC or target of all types of cells, e.g. of chondrocytes, osteoblasts and
CC fibroblasts. They can also be used: for treating pathological conditions
CC involving ISa10, such as damage to cartilage, trauma, rheumatoid
CC arthritis or osteoarthritis; for detecting the formation of cartilage
CC during embryonal development, physiological or therapeutic repair of
CC cartilage, or detecting regeneration of cartilage or chondrocytes during
CC transplantation of cartilage or chondrocytes; for selection and analysis
CC or for sorting, isolating or purification of chondrocytes and for in
CC vitro studies of differentiation of chondrocytes; and as a target for
CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or
CC other tissues where adhesion impairs the function of the tissue (all
CC claimed). ISa10 binding entities can be used to determine the
CC differentiation-state of cells during embryonic development, angiogenesis
CC or development of cancer, in pathological conditions such as rheumatoid
CC arthritis, osteoarthritis or cancer, in tissue regeneration or in
CC therapeutic and physiological repair of cartilage (claimed). A
CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also
CC claimed. ISa10 polynucleotides, vectors, host cells and methods of
CC producing recombinant ISa10 are also claimed

XX Sequence 1132 AA;

Query Match 100.0%; Score 1025; DB 3; Length 1132;
Best Local Similarity 100.0%; Pred. No. 2.4e-108;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLDP 60
DB 140 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLDP 199
QY 61 EQIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 120
DB 200 EQIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 259
QY 121 SHGGRPEAARLLVVVTDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
DB 260 SHGGRPEAARLLVVVTDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSSFL 319
QY 181 REIRTIASDPDERFFNV 198
DB 320 REIRTIASDPDERFFNV 337

RESULT 2
AAB64657
ID AAB64657 standard; protein; 1152 AA.

XX AAB64657;

XX 22-MAR-2001 (first entry)

XX Human secreted protein BLAST search protein SEQ ID NO: 167.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
XX vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO2000077197-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US014934.

XX 11-JUN-1999; 99US-0138599P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX Rosen CA, Ruben SM, Komatsoulis GA;
PI WPI; 2001-032312/04.

XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.

XX Disclosure; Page 543-546; 558pp; English.

XX The invention relates to the isolation of genes AAF32757-F32803 encoding
XX the human secreted proteins AAB64549-B64594. The sequence is a search
XX result from a BLASTX homology search. The genes and proteins are useful
XX for preventing, ameliorating or treating medical conditions, e.g. by
XX protein or gene therapy. The genes are isolated from a range of human
XX tissues disclosed in the specification. The nucleic acids, proteins,
XX antibodies and (ant)agonists are useful in the diagnosis, treatment and
XX prevention of: (a) cancer, e.g. breast and ovarian cancer, and other
XX cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal
XX tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
XX disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
XX rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders
XX such as myocardial ischaemias; (d) wound healing; (e) neurological
XX diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases
XX such as viral, bacterial, fungal and parasitic infections

XX Sequence 1152 AA;

Query Match 100.0%; Score 1025; DB 4; Length 1152;
Best Local Similarity 100.0%; Pred. No. 2.5e-108;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLDP 60
DB 140 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLDP 199
QY 61 EQIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 120
DB 200 EQIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 259
QY 121 SHGGRPEAARLLVVVTDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
DB 260 SHGGRPEAARLLVVVTDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSSFL 319
QY 181 REIRTIASDPDERFFNV 198
DB 320 REIRTIASDPDERFFNV 337

RESULT 3
AAB64658

ID AAB64658 standard; protein; 1152 AA.

XX AAB64658;

XX 22-MAR-2001 (first entry)

XX Human secreted protein BLAST search protein SEQ ID NO: 168.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
XX vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO2000077197-A1.

XX 21-DEC-2000.

```
XX 01-JUN-2000; 2000WO-US014934.
XX
XX 11-JUN-1999; 99US-0138599P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI; 2001-032312/04.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; Page 547-551; 558pp; English.
XX
XX The invention relates to the isolation of genes AAF32757-F32803 encoding
XX the human secreted proteins AAB64549-B64594. The sequence is used as a
XX query sequence for doing BLASTX searches to identify homologous
XX sequences. The genes and proteins are useful for preventing, ameliorating
XX or treating medical conditions, e.g. by protein or gene therapy. The
XX genes are isolated from a range of human tissues disclosed in the
XX specification. The nucleic acids, proteins, antibodies and (ant)agonists
XX are useful in the diagnosis, treatment and prevention of: (a) cancer,
XX e.g. breast and ovarian cancer, and other cancers of the adrenal gland,
XX bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
XX urogenital; (b) immune disorders e.g. Addison's disease, allergies,
XX autoimmune hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
XX Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
XX wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections
XX
XX SQ Sequence 1152 AA;

Query Match      100.0%; Score 1025; DB 4; Length 1152;
Best Local Similarity 100.0%; Pred. No. 2.5e-108;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GICARVDASFOQGS LAPTACRCTYMDVIVL DGSNIYPSWSEVQVFLRLVGLKLFIDP 60
   |||||
Db 140 GICARVDASFOQGS LAPTACRCTYMDVIVL DGSNIYPSWSEVQVFLRLVGLKLFIDP 199
   |||||

QY 61 EQIQVGLVQGESPVHEWSLGDFTKKEVVRAAKNLSRRREGRETKTQAQIMVACTEGFSQ 120
   |||||
Db 200 EQIQVGLVQGESPVHEWSLGDFTKKEVVRAAKNLSRRREGRETKTQAQIMVACTEGFSQ 259
   |||||

QY 121 SHGGRPEAARLLVVVT DGS HDGELPAAL KACAGRVTRYGIAVLGHYLRQRDPSSFL 180
   |||||
Db 260 SHGGRPEAARLLVVVT DGS HDGELPAAL KACAGRVTRYGIAVLGHYLRQRDPSSFL 319
   |||||

QY 181 REIRTIASDPDERFFENV 198
   |||||
Db 320 REIRTIASDPDERFFENV 337
   |||||

RESULT 4
AAV32242
ID AAY32242 standard; protein; 1167 AA.
XX
XX AC AAY32242;
XX
XX DT 15-FEB-2000 (first entry)
XX
XX DE Human integrin subunit alpha-10.
XX
XX Integrin alpha-10; Isal0; human; trauma; rheumatoid arthritis;
XX osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;
XX therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.
XX
XX OS Homo sapiens.

XX Key Location/Qualifiers
XX Peptide 1..22
XX /note= "signal peptide"
XX Protein 23..1145
XX /note= "mature protein"
XX Domain 23..1120
XX /note= "extracellular domain"
XX Modified-site 98
XX /note= "N-glycosylated"
XX Domain 162..359
XX /note= "I-domain"
XX Modified-site 336
XX /note= "N-glycosylated"
XX Modified-site 364
XX /note= "N-glycosylated"
XX Binding-site 494..502
XX /note= "cation binding site motif"
XX Binding-site 558..566
XX /note= "cation binding site motif"
XX Binding-site 620..628
XX /note= "cation binding site motif"
XX Modified-site 733
XX /note= "N-glycosylated"
XX Modified-site 839
XX /note= "N-glycosylated"
XX Modified-site 921
XX /note= "N-glycosylated"
XX Modified-site 1018
XX /note= "N-glycosylated"
XX Modified-site 1039
XX /note= "N-glycosylated"
XX Domain 1121..1145
XX /note= "transmembrane domain"
XX Domain 1122..1167
XX /note= "cytoplasmic domain, specifically claimed in Claim
XX 21"

XX WO9951639-A1.
XX
XX 14-OCT-1999.
XX
XX 31-MAR-1999; 99WO-SB0000544.
XX
XX 02-APR-1998; 98SE-00001164.
XX
XX 28-JAN-1999; 99SE-00000319.
XX
XX (ACTI-) ACTIVE BIOTECH AB.
XX
XX Lundgren-Akerlund E;
XX
XX WPI; 2000-052639/04.
XX
XX N-PSDB; AAZ34719.
XX
XX New isolated integrin subunit alpha-10, used as a marker or target
XX molecule for cells during development, regeneration and pathological
XX conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
XX inflammation.
XX
XX Claim 1; Fig 6; 90pp; English.
XX
XX This sequence represents novel human chondrocyte integrin subunit alpha-
XX 10 (Isal0). A splice variant is given in AAY32243. The invention relates
XX to a recombinant or isolated integrin heterodimer comprising the alpha10
XX subunit in association with subunit beta (especially beta-1). The
XX heterodimer and the subunit alpha-10 can be used as markers or targets of
XX all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
XX They can also be used: for treating pathological conditions involving
XX Isal0, such as damage to cartilage, trauma, rheumatoid arthritis or
XX osteoarthritis; for detecting the formation of cartilage during embryonal
XX development, physiological or therapeutic repair of cartilage, or
XX detecting regeneration of cartilage or chondrocytes during
XX transplantation of cartilage or chondrocytes; for selection and analysis
```

CC or for sorting, isolating or purification of chondrocytes and for in
CC vitro studies of differentiation of chondrocytes; and as a target for
CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or
CC other tissues where adhesion impairs the function of the tissue (all
CC claimed). ISa10 binding entities can be used to determine the
CC differentiation-state of cells during embryonic development, angiogenesis
CC or development of cancer, in pathological conditions such as rheumatoid
CC arthritis, osteoarthritis or cancer, in tissue regeneration or in
CC therapeutic and physiological repair of cartilage (claimed). A
CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also
CC claimed. ISa10 polynucleotides, vectors, host cells and methods of
CC producing recombinant ISa10 are also claimed

XX SQ Sequence 1167 AA;

Query Match 100.0%; Score 1025; DB 3; Length 1167;
Best Local Similarity 100.0%; Pred. No. 2.5e-108; Mismatches 0; Indels 0; Gaps 0;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFPQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLFIDP 60
Db 140 GICARVDASFPQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLFIDP 199
Qy 61 EQIQVGLVQGESPVHWSLGDPTKBEVVRAAKNLSRREGRETQTAQIMVACTEGFSQ 120
Db 200 EQIQVGLVQGESPVHWSLGDPTKBEVVRAAKNLSRREGRETQTAQIMVACTEGFSQ 259
Qy 121 SHGGRPEARLLVVVTGESHGDBELPAALKACEAGRVTRYGVIAVLGHYLRORDPSSFL 180
Db 260 SHGGRPEARLLVVVTGESHGDBELPAALKACEAGRVTRYGVIAVLGHYLRORDPSSFL 319
Qy 181 REIRTIASDPDERFFNV 198
Db 320 REIRTIASDPDERFFNV 337

RESULT 5

AAB64584
ID AAB64584 standard; protein; 1167 AA.

XX AC AAB64584;

XX DT 22-MAR-2001 (first entry)

XX DE Human secreted protein #37.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.

XX OS Homo sapiens.

XX XX WO200077197-A1.

XX XX 21-DEC-2000.

XX PF 01-JUN-2000; 2000WO-US014934.

XX PR 11-JUN-1999; 99US-0138599P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (ROSE/) ROSEN C A.

XX PI Rosen CA, Ruben SM, Komatsoulis GA;

XX DR WPI; 2001-032312/04.

XX DR N-PSDB; AAF32793.

XX PT Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.

PS Claim 11; Page 496-500; 558pp; English.

XX Sequences AAB64549-B64594 represent the amino acid sequences of 47 human
CC secreted proteins encoded by the genes AAF32757-F32803. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic
CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)
CC immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
XX infections

XX SQ Sequence 1167 AA;

Query Match 100.0%; Score 1025; DB 4; Length 1167;
Best Local Similarity 100.0%; Pred. No. 2.5e-108; Mismatches 0; Indels 0; Gaps 0;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFPQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLFIDP 60
Db 140 GICARVDASFPQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLFIDP 199
Qy 61 EQIQVGLVQGESPVHWSLGDPTKBEVVRAAKNLSRREGRETQTAQIMVACTEGFSQ 120
Db 200 EQIQVGLVQGESPVHWSLGDPTKBEVVRAAKNLSRREGRETQTAQIMVACTEGFSQ 259
Qy 121 SHGGRPEARLLVVVTGESHGDBELPAALKACEAGRVTRYGVIAVLGHYLRORDPSSFL 180
Db 260 SHGGRPEARLLVVVTGESHGDBELPAALKACEAGRVTRYGVIAVLGHYLRORDPSSFL 319
Qy 181 REIRTIASDPDERFFNV 198
Db 320 REIRTIASDPDERFFNV 337

RESULT 6

ABP99490
ID ABP99490 standard; protein; 1167 AA.

XX AC ABP99490;

XX DT 26-MAR-2003 (first entry)

XX DE Human secreted protein SEQ ID NO 434.

XX KW Human; secreted protein; nootropic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; antinflammatory; anti-HIV;
KW vulnery; antibacterial; antiparkinsonian; antiscikling; antianemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
XX gene therapy.

XX OS Homo sapiens.

XX XX WO200277186-A2.

XX PN 03-OCT-2002.

XX PD 26-MAR-2002; 2002WO-US009188.

XX PF 27-MAR-2001; 2001US-0278650P.

XX PR 12-SEP-2001; 2001US-00950082.

XX PR 12-SEP-2001; 2001US-00950083.

XX XX

(HUMA-) HUMAN GENOME SCI INC.
 PA Reen CA, Ruben SM;
 PI WPI; 2003-040583/03.
 DR N-PSDB; ABZ66911.
 XX
 XX New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
 PT West Nile fever.
 XX
 XX Claim 1; Page 1395-1398; 2423pp; English.
 PS
 XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections
 XX
 XX Sequence 1167 AA;
 SQ
 Query Match 100.0%; Score 1025; DB 6; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 2.5e-108;
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYVPWSVQTFRLRLVGLKLFIDP 60
 DB 140 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYVPWSVQTFRLRLVGLKLFIDP 199
 QY 61 EQIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSRREGRETAKTAQAINVACTEGFSQ 120
 DB 200 EQIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSRREGRETAKTAQAINVACTEGFSQ 259
 QY 121 SHGGRPEARLLVVVTDGESHGDELPALAKACEAGRVTTRYGIAVLGHYLRQRDPSSFL 180
 DB 260 SHGGRPEARLLVVVTDGESHGDELPALAKACEAGRVTTRYGIAVLGHYLRQRDPSSFL 319
 QY 181 REIRTIASDPDERFFNV 198
 DB 320 REIRTIASDPDERFFNV 337
 RESULT 7
 ABR00964
 ID ABR00964 standard; protein; 1167 AA.
 XX
 AC ABR00964;
 XX
 XX 12-MAY-2003 (first entry)
 DT
 XX Human gene 18-encoded secreted protein HA0AG15, SEQ ID NO:445.
 DE
 XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnery; chromosome 1q21.
 XX
 OS Homo sapiens.
 XX
 XX WO200277013-A2.
 PN
 03-OCT-2002.
 XX
 XX 26-MAR-2002; 2002WO-US009370.
 XX
 XX 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI WPI; 2003-040578/03.
 DR N-PSDB; ABZ73298.
 XX
 XX New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX
 XX Claim 13; Page 1387-1390; 2474pp; English.
 PS
 XX ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein of the invention
 XX
 XX Sequence 1167 AA;
 SQ
 Query Match 100.0%; Score 1025; DB 6; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 2.5e-108;
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYVPWSVQTFRLRLVGLKLFIDP 60
 DB 140 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYVPWSVQTFRLRLVGLKLFIDP 199
 QY 61 EQIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSRREGRETAKTAQAINVACTEGFSQ 120
 DB 200 EQIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSRREGRETAKTAQAINVACTEGFSQ 259
 QY 121 SHGGRPEARLLVVVTDGESHGDELPALAKACEAGRVTTRYGIAVLGHYLRQRDPSSFL 180
 DB 260 SHGGRPEARLLVVVTDGESHGDELPALAKACEAGRVTTRYGIAVLGHYLRQRDPSSFL 319
 QY 181 REIRTIASDPDERFFNV 198
 DB 320 REIRTIASDPDERFFNV 337
 RESULT 8
 ADA44026
 ID ADA44026 standard; protein; 1167 AA.
 XX
 XX ADA44026;
 AC
 XX

DT 20-NOV-2003 (first entry)
 XX Human secreted protein SEQ ID 214.
 DE
 XX
 KW Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;
 KW Neuroprotective; Cerebroprotective; Antianemic.
 XX
 OS Homo sapiens.
 XX
 XX WO2003000865-A2.
 PN
 XX 03-JAN-2003.
 PD
 XX 26-MAR-2002; 2002WO-US009105.
 PF
 XX 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2003-184045/18.
 DR
 XX N-PSDB; ADA43832.
 XX
 XX A human secreted protein and nucleic acids useful for preparing a
 PT diagnostic or pharmaceutical composition for diagnosing or treating
 PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,
 PT retinopathy, neuropathy.
 XX
 PS Claim 1; SEQ ID NO 214; 701pp; English.
 XX
 CC The invention relates to novel genes and their fragments which are useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids and proteins
 CC are useful in the diagnosis, treatment and prevention of conditions
 CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy.
 CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,
 CC infection, cataract, renal disorders, or endocrine disorders. The present
 CC sequence was used to illustrate the invention.
 XX
 SQ Sequence 1167 AA;
 Query Match 100.0%; Score 1025; DB 6; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 2.5e-108; Indels 0; Gaps 0;
 Matches 198; Conservative 0; Mismatches 0;
 Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 60
 Db 140 GICARVDASFOQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 199
 Qy 61 EQIQVGLVQVGS PVEHWSLGDFTKBEVVRAAKNLSRREGRETKTQAQIMVACTEGFSQ 120
 Db 200 EQIQVGLVQVGS PVEHWSLGDFTKBEVVRAAKNLSRREGRETKTQAQIMVACTEGFSQ 259
 Qy 121 SHGGRPBAARLLVVVTGESHGDELPAAKACAGRVTRYGVIAVLGHYLRORDPSSFL 180
 Db 260 SHGGRPBAARLLVVVTGESHGDELPAAKACAGRVTRYGVIAVLGHYLRORDPSSFL 319
 Qy 181 REIRTIASDDPDERFFNV 198
 Db 320 REIRTIASDDPDERFFNV 337
 RESULT 9
 ADQ19290
 ID ADQ19290 standard; protein; 1167 AA.
 XX
 AC ADQ19290;
 XX
 DT 26-AUG-2004 (first entry)
 XX

XX Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.
 DE
 XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
 KW
 OS Homo sapiens.
 XX
 XX WO2004048938-A2.
 PN
 XX 10-JUN-2004.
 PD
 XX 26-NOV-2003; 2003WO-US038193.
 PF
 XX 26-NOV-2002; 2002US-0429739P.
 PR
 XX (PROT-) PROTEIN DESIGN LABS INC.
 PA
 XX Aziz N, Ginsburg WM, Zlotnik A;
 PI
 XX WPI; 2004-441208/41.
 DR
 XX
 XX Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.
 XX
 PS Example 2; SEQ ID NO 2109; 210pp; English.
 XX
 CC The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC protein of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 SQ Sequence 1167 AA;
 Query Match 100.0%; Score 1025; DB 8; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 2.5e-108; Indels 0; Gaps 0;
 Matches 198; Conservative 0; Mismatches 0;
 Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 60
 Db 140 GICARVDASFOQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 199
 Qy 61 EQIQVGLVQVGS PVEHWSLGDFTKBEVVRAAKNLSRREGRETKTQAQIMVACTEGFSQ 120
 Db 200 EQIQVGLVQVGS PVEHWSLGDFTKBEVVRAAKNLSRREGRETKTQAQIMVACTEGFSQ 259
 Qy 121 SHGGRPBAARLLVVVTGESHGDELPAAKACAGRVTRYGVIAVLGHYLRORDPSSFL 180
 Db 260 SHGGRPBAARLLVVVTGESHGDELPAAKACAGRVTRYGVIAVLGHYLRORDPSSFL 319
 Qy 181 REIRTIASDDPDERFFNV 198
 Db 320 REIRTIASDDPDERFFNV 337
 RESULT 10
 AAU76862
 ID AAU76862 standard; protein; 195 AA.
 XX
 AC AAU76862;
 XX
 DT 21-MAY-2002 (first entry)
 XX

DE XX Human integrin alpha subunit Alpha 10 variant A domain.

KW Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;
KW mitein.

XX
XX Homo sapiens.
OS Synthetic.

XX
XX Key Location/Qualifiers
FH Misc-difference 193
FT /note= "Wild-type Ile substituted by any other amino
FT acid"
XX
XX WO200209737-A1.
XX
XX 07-FEB-2002.
XX
XX 31-JUL-2001; 2001WO-US023957.
XX
XX 31-JUL-2000; 2000US-0221950P.
PR 11-JAN-2001; 2001US-00758493.
PR 13-MAR-2001; 2001US-00805354.
XX
XX (GEO) GEN HOSPITAL CORP.
PA
XX Arnaout AM, Li R, Xiong J;
PI WPI; 2002-188687/24.
DR
XX Novel high affinity integrin polypeptide useful for treating restenosis
PT and parasitic diseases, comprises all or part of variant integrin alpha
PT subunit A domain or variant integrin beta subunit A-like domain.
XX
XX Claim 53; Page: 55pp; English.

XX The invention relates to a high affinity integrin polypeptide comprising
CC all or part of a variant integrin alpha subunit A domain or a variant
CC integrin beta subunit A-like domain. The polypeptide, preferably the
CC Cdlb alpha subunit A domain, where I at residue 332 has been replaced by
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
CC V at residue 315 and A at residue 320 have been replaced by C, is useful
CC for determining if a test compound is a candidate compound for binding to
CC Cdlb or for treating an inflammatory disorder, by contacting a test
CC compound with the polypeptide and determining if the test compound binds
CC to the polypeptide. The integrin subunits are useful for reducing
CC skeletal muscle injury, for treating disorders caused by ischaemia-
CC reperfusion injury, immune complexes, restenosis and parasitic diseases,
CC to purify variant integrin polypeptide ligands and as bait proteins in
CC two-hybrid or three-hybrid assays. This sequence represents a human
CC integrin alpha subunit Alpha 10 variant A domain. Note: This variant
CC sequence is not featured in the specification but has been derived from
CC the wild-type protein shown in AAU76853

XX
SQ Sequence 195 AA;

Query Match 89.0%; Score 912; DB 5; Length 195;
Best Local Similarity 100.0%; Pred. No. 2e-96;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDPEQIQVGLVQYGESPVHWSLGD 82
Db 1 CPTMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDPEQIQVGLVQYGESPVHWSLGD 60

Qy 83 FRTKEEVVRAAKNLSRREGRTKTAQAIMVACTEGFSQSHGGRPEARLLVVVTDGSHD 142
Db 61 FRTKEEVVRAAKNLSRREGRTKTAQAIMVACTEGFSQSHGGRPEARLLVVVTDGSHD 120

Qy 143 GEELPAALKACEAGRVTRYGIAVLGHVLRQRDPSSFLREIRTTASPDPRFFENV 198
Db 121 GEELPAALKACEAGRVTRYGIAVLGHVLRQRDPSSFLREIRTTASPDPRFFENV 176

RESULT 11
AAU76853
ID AAU76853 standard; protein; 195 AA.
XX
XX AAU76853;
XX
XX 21-MAY-2002 (first entry)
XX
XX Human integrin alpha subunit Alpha 10 A domain.
XX
XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
KW ischaemia-reperfusion injury; immune complex; parasitic disease;
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.
XX
XX Homo sapiens.
OS
XX WO200209737-A1.
PN
XX 07-FEB-2002.
XX
XX 31-JUL-2001; 2001WO-US023957.
XX
XX 31-JUL-2000; 2000US-0221950P.
PR 11-JAN-2001; 2001US-00758493.
PR 13-MAR-2001; 2001US-00805354.
XX
XX (GEO) GEN HOSPITAL CORP.
PA
XX Arnaout AM, Li R, Xiong J;
PI WPI; 2002-188687/24.
DR
XX Novel high affinity integrin polypeptide useful for treating restenosis
PT and parasitic diseases, comprises all or part of variant integrin alpha
PT subunit A domain or variant integrin beta subunit A-like domain.
XX
XX Example 2; Fig 5; 55pp; English.

XX The invention relates to a high affinity integrin polypeptide comprising
CC all or part of a variant integrin alpha subunit A domain or a variant
CC integrin beta subunit A-like domain. The polypeptide, preferably the
CC Cdlb alpha subunit A domain, where I at residue 332 has been replaced by
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
CC V at residue 315 and A at residue 320 have been replaced by C, is useful
CC for determining if a test compound is a candidate compound for binding to
CC Cdlb or for treating an inflammatory disorder, by contacting a test
CC compound with the polypeptide and determining if the test compound binds
CC to the polypeptide. The integrin subunits are useful for reducing
CC skeletal muscle injury, for treating disorders caused by ischaemia-
CC reperfusion injury, immune complexes, restenosis and parasitic diseases,
CC to purify variant integrin polypeptide ligands and as bait proteins in
CC two-hybrid or three-hybrid assays. This sequence represents the human
CC integrin alpha subunit Alpha 10 A domain

XX
SQ Sequence 195 AA;

Query Match 89.0%; Score 912; DB 5; Length 195;
Best Local Similarity 100.0%; Pred. No. 2e-96;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDPEQIQVGLVQYGESPVHWSLGD 82
Db 1 CPTMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDPEQIQVGLVQYGESPVHWSLGD 60

Qy 83 FRTKEEVVRAAKNLSRREGRTKTAQAIMVACTEGFSQSHGGRPEARLLVVVTDGSHD 142
Db 61 FRTKEEVVRAAKNLSRREGRTKTAQAIMVACTEGFSQSHGGRPEARLLVVVTDGSHD 120

Qy 143 GEELPAALKACEAGRVTRYGIAVLGHVLRQRDPSSFLREIRTTASPDPRFFENV 198
Db 121 GEELPAALKACEAGRVTRYGIAVLGHVLRQRDPSSFLREIRTTASPDPRFFENV 176

Db 121 GEELPAALKACEAGRVTRYGIVLGHYLRQRDPSSFLREIRTIASDPDRFFPNV 176

RESULT 12

ADR23227

ID ADR23227 standard; protein; 198 AA.

XX

AC ADR23227;

XX

DT 04-NOV-2004 (first entry)

XX

DE Human integrin alpha 10 subunit A domain.

XX

KW Human; integrin; inflammation; antiinflammatory; vasotropic.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT Domain 1..195

FT /label = A_domain

FT Misc-difference 193

FT /note= "Invariant Ile residue"

XX

PN WO2004066914-A2.

XX

PD 12-AUG-2004.

XX

PP 12-MAY-2003; 2003WO-US014919.

XX

PR 10-MAY-2002; 2002US-00144259.

XX

PA (GSHO) GEN HOSPITAL CORP.

XX

PI Arnaut MA, Li R, Xiong J;

XX

DR WPI; 2004-593980/57.

DR GENBANK; XP_002097.

XX

PT Novel variant integrin CD11b alpha subunit polypeptide, useful for

PT determining candidate compound for binding to CD11b, and for determining

PT compound as activation-dependent ligand.

XX

PS Disclosure; SEQ ID NO 7; 128pp; English.

XX

CC The present sequence is that of the human integrin alpha 10 subunit A

CC domain. This includes an invariant Ile residue. The invention features

CC variant integrin alpha subunit polypeptides in which the invariant Ile is

CC substituted by Gly, Ala or some other amino acid (e.g. Val) or is

CC deleted. The polypeptide can include part or all of the A domain.

CC Replacing the invariant Ile creates a variant integrin polypeptide that

CC is more active (i.e. in solution has a greater proportion of ligand-

CC forming polypeptides) than the wild-type form of the subunit. Variant

CC integrin polypeptides of the invention are useful in assays for compounds

CC that bind to a variant ligand, that interfere with or enhance the binding

CC of an integrin ligand to integrin, and for identifying antibodies-

CC specific ligands. They are also useful for generating antibodies, e.g.

CC monoclonal antibodies, which bind to the high efficiency form of an

CC integrin. Some such antibodies recognise an epitope that is either not

CC present or not accessible on an integrin that is in a lower affinity

CC conformation. The invention also provides methods of administering a

CC variant integrin polypeptide, or an antibody that selectivity binds it,

CC to identify a ligand which binds to an active integrin. Such assays are

CC useful for diagnosing inflammation, e.g. occult inflammation (e.g.

CC abscess or an active arteriosclerotic lesion). Variant integrin

CC polypeptides can also be used to affect the bioavailability of a variant

CC integrin polypeptide ligand and to treat disorders associated with

CC aberrant or unwanted integrin expression or activity, such as vascular

CC injury.

XX

SQ Sequence 198 AA;

Query Match 89.0%; Score 912; DB 8; Length 198;

Best Local Similarity 100.0%; Pred. No. 2.1e-96;

Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 CPTYMDVIVLDGNSIYPMSSEVQTFLLRLVGLKFLIDPEIQVGLVQYGSPVHWSLGD 82

DB 1 CPTYMDVIVLDGNSIYPMSSEVQTFLLRLVGLKFLIDPEIQVGLVQYGSPVHWSLGD 60

QY 83 FRTKEEVVRAAKNLSRRREGRETAKAIVVACTEGFSQSHGSRPEAARLLVVVTDGESH 142

DB 61 FRTKEEVVRAAKNLSRRREGRETAKAIVVACTEGFSQSHGSRPEAARLLVVVTDGESH 120

QY 143 GEELPAALKACEAGRVTRYGIVLGHYLRQRDPSSFLREIRTIASDPDRFFPNV 198

DB 121 GEELPAALKACEAGRVTRYGIVLGHYLRQRDPSSFLREIRTIASDPDRFFPNV 176

RESULT 13

ADR41424

ID ADR41424 standard; protein; 1049 AA.

XX

AC ADR41424;

XX

DT 07-OCT-2004 (first entry)

XX

DE Human CD-like molecule HA0AD02, SEQ ID NO:223.

XX

KW Human; CD-like molecule; cluster of differentiation; diagnosis;

KW prevention; immune disorder; immunodeficiency; autoimmune disorder;

KW blood-related disorder; haematological disorder; haemostatic disorder;

KW thrombolytic disorder; hyperproliferative disorder; cancer; tumour;

KW apoptotic disorder; cardiovascular disorder; respiratory disorder;

KW angiogenic disorder; neovascularisation; neurological disorder;

KW endocrine disorder; reproductive system disorder; infectious disease;

KW gastrointestinal disorder; drug screening; tissue regeneration;

KW chemotaxis; gene therapy; antibody therapy; drug targeting;

KW chromosome mapping; forensic analysis; immunophenotyping; cytostatic;

KW haemostatic; tranquiliser; vulnery; antinflammatory; nephrotropic;

KW cardiant; antiallergic; anti-HIV; antirheumatic; antiarthritic;

KW antipsoratic; immunosuppressive; vasotropic; neurotropic; neuroprotective;

KW antithyroid; thyromimetic; gynaecological; virucide; hepatotropic;

KW antibacterial; dermatological; chromosome 1q21.

XX

OS Homo sapiens.

XX

PN WO200226930-A2.

XX

PD 04-APR-2002.

XX

PP 25-SEP-2001; 2001WO-US029838.

XX

PR 26-SEP-2000; 2000US-0235484P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Birse CB;

XX

DR WPI; 2002-405050/43.

DR N-PSDB; ADR41248.

XX

PT Novel polynucleotides and polypeptides useful for treating, preventing or

PT ameliorating cardiovascular, renal, neurovascular, and autoimmune

PT disorders.

XX

PS Claim 11; SEQ ID NO 223; 1243pp; English.

XX

CC The invention relates to 167 novel human CD (cluster of differentiation)-

CC like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid:11)-

XX

SQ Sequence 1049 AA;

Query Match 82.0%; Score 840.5; DB 5; Length 1049;

Best Local Similarity 86.6%; Pred. No. 4.3e-87;

Matches 168; Conservative 5; Mismatches 14; Indels 7; Gaps 2;


```
Db 320 REIRTIASDPDRPFNFV 337
|||||
RESULT 2
US-10-741-601-532
; Sequence 532, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C0001500
; CURRENT FILING DATE: 2003-12-22
; CURRENT APPLICATION NUMBER: US/10/741.601
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-532

Query Match 100.0%; Score 1025; DB 16; Length 1177;
Best Local Similarity 100.0%; Pred. No. 1.9e-100;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASPOQSLAPTAQRCPTQYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 60
Db 140 GICARVDASPOQSLAPTAQRCPTQYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 199
Qy 61 EQIQVGLVQGESPVHWSLGDFTKBEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 120
Db 200 EQIQVGLVQGESPVHWSLGDFTKBEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 259
Qy 121 SHGGRPAARLLVVVTGESHGDBELPAALKACBAGRVTTRYGIAVLGHYLRQRDPSSFL 180
Db 260 SHGGRPAARLLVVVTGESHGDBELPAALKACBAGRVTTRYGIAVLGHYLRQRDPSSFL 319
Qy 181 REIRTIASDPDRPFNFV 198
Db 320 REIRTIASDPDRPFNFV 337

RESULT 3
US-09-805-354-7
; Sequence 7, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805.354
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-7

Query Match 89.0%; Score 912; DB 10; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.1e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQGESPVHWSLGD 82
Db 1 CPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQGESPVHWSLGD 60
Qy 83 FRTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESH 142
Db 61 FRTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESH 120
Qy 143 GEELPAALKACEAGRVTTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFV 198
Db 121 GEELPAALKACEAGRVTTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFV 176

Query Match 89.0%; Score 912; DB 11; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.1e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQGESPVHWSLGD 82
Db 1 CPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQGESPVHWSLGD 60
Qy 83 FRTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESH 142
Db 61 FRTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESH 120
Qy 143 GEELPAALKACEAGRVTTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFV 198
Db 121 GEELPAALKACEAGRVTTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFV 176

RESULT 5
US-10-144-259-7
; Sequence 7, Application US/10144259
; Publication No. US20030109691A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-548001
; CURRENT APPLICATION NUMBER: US/10/144,259
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 195
; TYPE: PRT
```

```
Db 1 CPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQGESPVHWSLGD 60
Qy 83 FRTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESH 142
Db 61 FRTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESH 120
Qy 143 GEELPAALKACEAGRVTTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFV 198
Db 121 GEELPAALKACEAGRVTTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFV 176

RESULT 4
US-09-758-493-7
; Sequence 7, Application US/09758493
; Publication No. US20040086935A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 00786-804001
; CURRENT APPLICATION NUMBER: US/09/758,493
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-758-493-7

Query Match 89.0%; Score 912; DB 11; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.1e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQGESPVHWSLGD 82
Db 1 CPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQGESPVHWSLGD 60
Qy 83 FRTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESH 142
Db 61 FRTKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESH 120
Qy 143 GEELPAALKACEAGRVTTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFV 198
Db 121 GEELPAALKACEAGRVTTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFV 176

RESULT 5
US-10-144-259-7
; Sequence 7, Application US/10144259
; Publication No. US20030109691A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-548001
; CURRENT APPLICATION NUMBER: US/10/144,259
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 195
; TYPE: PRT
```

```

; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-540-313

Query Match      53.1%; Score 544; DB 14; Length 707;
Best Local Similarity 53.0%; Pred. No. 4e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1

QY      1  GICARVDASFPQGS LAPTAQRCTPTVMVIVLDGNSIYPWSEVQTFLRLVLGKLFIDP 60
DB      153  GMSCRVNSNFRFSKTVAPALQRCQTYMDIIVIVLDGNSIYPWSEVQHFLLNLIKFFYIGP 212

QY      61  EQIQVGLVYGESFPVHWSLGDRTKEEVVRAAKNLSRRREGRETAKTAQAIWVACTEGFSQ 120
DB      213  GQIQVGVQYGEDVHVEFLHNDYRSVKDVVEAASHIEQRGTETRTAFGIFARSEAFQK 272

QY      121  SHGGRPEARLLVVVTDGESHGDEELPAALKACBAGRVTTRYGIAVLGHYLRQRDPDSFL 180
DB      273  --GGRKCAKVMIVITDGHSDPDLEKVIQQSERDENVTRVAVAVLGYNNRRGINPETFL 330

QY      181  REIRTIASDPDERFFNV 198
DB      331  NEIKYIASDPDDKHFNV 348

RESULT 8
US-09-984-130-43
; Sequence 43, Application US/09984130
; Publication No. US2003005231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-43

Query Match      53.1%; Score 544; DB 10; Length 1034;
Best Local Similarity 53.0%; Pred. No. 6.8e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1

QY      1  GICARVDASFPQGS LAPTAQRCTPTVMVIVLDGNSIYPWSEVQTFLRLVLGKLFIDP 60
DB      137  GMSCRVNSNFRFSKTVAPALQRCQTYMDIIVIVLDGNSIYPWSEVQHFLLNLIKFFYIGP 196

QY      61  EQIQVGLVYGESFPVHWSLGDRTKEEVVRAAKNLSRRREGRETAKTAQAIWVACTEGFSQ 120
DB      197  GQIQVGVQYGEDVHVEFLHNDYRSVKDVVEAASHIEQRGTETRTAFGIFARSEAFQK 256

QY      121  SHGGRPEARLLVVVTDGESHGDEELPAALKACBAGRVTTRYGIAVLGHYLRQRDPDSFL 180
DB      257  --GGRKCAKVMIVITDGHSDPDLEKVIQQSERDENVTRVAVAVLGYNNRRGINPETFL 314

QY      181  REIRTIASDPDERFFNV 198

```

```
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 9
; Sequence 43, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-43

Query Match 53.1%; Score 544; DB 10; Length 1034;
Best Local Similarity 53.0%; Pred. No. 6.8e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 137 GMSRVNSNFRFSKTVPALQRCQTYMDIVIVLDGNSIYPWSEVQHFLLNLIKFFVIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDFTKBEVRAAKNLSRREGRTKTAQAIMVACTEGFSQ 120
Db 197 GOIQGVVQYGEDVVFHFLNDYRSVKDVVEAAASHIEQRGGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRGAKKVMIVITDGHSDSPDLEKVIQOESRDNVTRYAVAVLGYNNRRGINPTFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 11
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 810
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-810

Query Match 53.1%; Score 544; DB 15; Length 1188;
Best Local Similarity 53.0%; Pred. No. 8.2e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 137 GMSRVNSNFRFSKTVPALQRCQTYMDIVIVLDGNSIYPWSEVQHFLLNLIKFFVIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDFTKBEVRAAKNLSRREGRTKTAQAIMVACTEGFSQ 120
Db 197 GOIQGVVQYGEDVVFHFLNDYRSVKDVVEAAASHIEQRGGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRGAKKVMIVITDGHSDSPDLEKVIQOESRDNVTRYAVAVLGYNNRRGINPTFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 12
US-09-984-130-35

Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 10
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338

Query Match 53.1%; Score 544; DB 10; Length 1034;
Best Local Similarity 53.0%; Pred. No. 6.8e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 137 GMSRVNSNFRFSKTVPALQRCQTYMDIVIVLDGNSIYPWSEVQHFLLNLIKFFVIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDFTKBEVRAAKNLSRREGRTKTAQAIMVACTEGFSQ 120
Db 197 GOIQGVVQYGEDVVFHFLNDYRSVKDVVEAAASHIEQRGGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRGAKKVMIVITDGHSDSPDLEKVIQOESRDNVTRYAVAVLGYNNRRGINPTFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332
```

Query Match	53.1%;	Score 544;	DB 10;	Length 1189;
Best Local Similarity	53.0%;	Pred. No. 8.2e-49;		

APPLICANT: Anderson, David W.,	US/10/01-01
APPLICANT: Boldog, Ferenc,	US/10/01-01
APPLICANT: Burgess, Catherine,	US/10/01-01
APPLICANT: Catterton, Elina,	US/10/01-01
APPLICANT: Edinger, Shlomit,	US/10/01-01
APPLICANT: Ellerman, Karen,	US/10/01-01
APPLICANT: Gerlach, Valerie,	US/10/01-01
APPLICANT: Gorman, Linda,	US/10/01-01
APPLICANT: Guo, Xiaojia,	US/10/01-01
APPLICANT: Ji, Weizhen,	US/10/01-01
APPLICANT: Kekuda, Ramesh,	US/10/01-01
APPLICANT: Leach, Martin,	US/10/01-01
APPLICANT: Li, Li,	US/10/01-01
APPLICANT: Miller, Charles,	US/10/01-01
APPLICANT: Patturajan, Meera,	US/10/01-01
APPLICANT: Reiger, Daniel,	US/10/01-01
APPLICANT: Rothenberg, Mark,	US/10/01-01
APPLICANT: Shimkets, Richard,	US/10/01-01
APPLICANT: Smithson, Glenda,	US/10/01-01
APPLICANT: Szytker, Kimberly,	US/10/01-01
APPLICANT: Taupier, Raymond, jr.,	US/10/01-01
APPLICANT: Vernet, Corine,	US/10/01-01
APPLICANT: Voss, Edward,	US/10/01-01
APPLICANT: Zernhusen, Brian,	US/10/01-01
APPLICANT: Zhong, Mei	US/10/01-01
TITLE OF INVENTION: THERAPEUTIC PO	US/10/01-01
FILE REFERENCE: 21402-462A	US/10/01-01
CURRENT APPLICATION NUMBER: US/10/01-01	US/10/01-01
CURRENT FILING DATE: 2002-10-01	US/10/01-01
PRIOR APPLICATION NUMBER: 60/326,4	US/10/01-01
PRIOR FILING DATE: 2001-10-02	US/10/01-01
PRIOR APPLICATION NUMBER: 60/327,9	US/10/01-01
PRIOR FILING DATE: 2001-10-09	US/10/01-01
PRIOR APPLICATION NUMBER: 60/328,0	US/10/01-01
PRIOR FILING DATE: 2001-10-09	US/10/01-01
PRIOR APPLICATION NUMBER: 60/328,0	US/10/01-01
PRIOR FILING DATE: 2001-10-09	US/10/01-01
PRIOR APPLICATION NUMBER: 60/381,1	US/10/01-01
PRIOR FILING DATE: 2002-05-16	US/10/01-01
PRIOR APPLICATION NUMBER: 60/371,9	US/10/01-01
PRIOR FILING DATE: 2002-04-12	US/10/01-01
PRIOR APPLICATION NUMBER: 60/327,3	US/10/01-01
PRIOR FILING DATE: 2001-10-05	US/10/01-01
PRIOR APPLICATION NUMBER: 60/328,0	US/10/01-01
PRIOR FILING DATE: 2001-10-09	US/10/01-01
PRIOR APPLICATION NUMBER: 60/374,7	US/10/01-01
PRIOR FILING DATE: 2001-10-12	US/10/01-01

```
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 4
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-4

Query Match      53.1%; Score 544; DB 15; Length 1189;
Best Local Similarity 53.0%; Pred. No. 8.2e-45;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFPQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 60
Db 137 GMSRVNSNFRFSKTVAPALQRCQTYMDIVLDGNSIYPWVEVQHFLINILKIFYIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDPRTKKEEVVRAAKNLSRREGRETCTAAQIMVACTEGFSQ 120
Db 197 GQIQVGVVQYGEDVVFHFLNDYRSVKDVVEAASHIEQRGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTGESHGDEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRGAKKVMIVITDGESHSDPLEKVIQQSERDNTVRYAVAVLGYNNRGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 15
US-10-108-260A-3386
; Sequence 3386, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3386
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3386

Query Match      52.8%; Score 541; DB 15; Length 437;
Best Local Similarity 52.5%; Pred. No. 4.3e-49;
Matches 104; Conservative 39; Mismatches 53; Indels 2; Gaps 1;

Qy 1 GICARVDASFPQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 60
Db 137 GMSRVNSNFRFSKTVAPALQRCQTYMDIVLDGNSIYPWVEVQHFLINILKIFYIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDPRTKKEEVVRAAKNLSRREGRETCTAAQIMVACTEGFSQ 120
Db 197 GQIQVGVVQYGEDVVFHFLNDYRSVKDVVEAASHIEQRGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTGESHGDEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRGAKKVMIVITDGESHSDPLEKVIQQSERDNTVRYAVAVLGYNNRGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332
```

Search completed: April 6, 2005, 12:53:02
Job time : 25.7306 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 8.45027 Seconds
(without alignments)
1749.117 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025

Sequence: 1 GICARVDASFPQGS LAPTA.....FLREIRTIASDPDRFFNV 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pap:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pap:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pap:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pap:*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pap:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	544	53.1	1217	4	US-09-949-016-7892
2	465.5	45.4	1180	4	US-09-000-004A-2
3	452.5	44.1	214	4	US-10-061-658-5
4	449.5	43.9	214	4	US-10-061-658-6
5	449.5	43.9	214	4	US-10-061-658-9
6	441	43.0	1181	4	US-09-000-004A-4
7	441	43.0	1181	4	US-09-949-016-6189
8	441	43.0	1195	4	US-09-949-016-10747
9	420.5	41.0	1183	4	US-09-532-310B-6
10	396	38.6	1183	4	US-09-532-310B-5
11	205	20.0	496	1	US-08-463-128-37
12	205	20.0	496	1	US-08-463-180-37
13	205	20.0	496	2	US-08-001-078A-1
14	205	20.0	496	2	US-08-897-443-4
15	205	20.0	496	2	US-08-463-218-1
16	205	20.0	496	4	US-09-949-016-11306
17	205	20.0	496	5	PCT-US94-00253-1
18	198	19.3	413	1	US-08-485-618-101
19	198	19.3	413	2	US-08-605-672-101
20	198	19.3	413	2	US-08-482-293A-101
21	198	19.3	413	2	US-08-943-363-101
22	198	19.3	413	3	US-09-193-043-101
23	198	19.3	413	4	US-09-688-307A-101
24	198	19.3	413	4	US-09-350-259-101
25	196	19.1	956	2	US-08-897-443-3
26	191	18.6	638	2	US-08-897-443-1
27	191	18.6	915	4	US-09-907-794A-34

28	191	18.6	915	4	US-09-905-125A-34	Sequence 34, Appl
29	191	18.6	915	4	US-09-902-775A-34	Sequence 34, Appl
30	191	18.6	915	4	US-09-906-700-34	Sequence 34, Appl
31	191	18.6	915	4	US-09-903-603A-34	Sequence 34, Appl
32	191	18.6	915	4	US-09-904-920A-34	Sequence 34, Appl
33	191	18.6	915	4	US-09-909-064-34	Sequence 34, Appl
34	191	18.6	915	4	US-09-905-381A-34	Sequence 34, Appl
35	191	18.6	915	4	US-09-906-618-34	Sequence 34, Appl
36	191	18.6	956	4	US-09-949-016-6215	Sequence 6215, Ap
37	191	18.6	963	4	US-09-949-016-11519	Sequence 11519, A
38	191	18.6	963	4	US-09-949-016-11520	Sequence 11520, A
39	185.5	18.1	584	4	US-09-949-016-10340	Sequence 10340, A
40	185.5	18.1	584	4	US-09-949-016-10341	Sequence 10341, A
41	177.5	17.3	1178	1	US-08-199-776-2	Sequence 2, Appli
42	177.5	17.3	1178	3	US-08-683-731-2	Sequence 2, Appli
43	177.5	17.3	1178	3	US-08-879-338-2	Sequence 2, Appli
44	177.5	17.3	1178	5	PCT-US95-02044-2	Sequence 2, Appli
45	177.5	17.3	1179	4	US-09-293-238B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-949-016-7892
; Sequence 7892, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7892
; LENGTH: 1217
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-7892

Query Match 53.1%; Score 544; DB 4; Length 1217;
Best Local Similarity 53.0%; Pred. No. 3.9e-59;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

QY	1	GICARVDASFPQGS LAPTAQRCPTYMDVIVLDGNSIYPWSEVQTFELRLVGLFIDP	60
DB	166	GMCSRVSNNFRFSKTVAPALQRCQTDIVIVLDGNSIYPWSEVQHFILNLIKFFYGP	225
QY	61	EQIQVGLVQYGESPVHWSLGDFTKEBVRRAKNLSRREGTKTAQAINVACTEGFSQ	120
DB	226	GOIQGVVQYGEDVVHFEHLNDYRSVKDVEAASHIEQRCGTETRTAFGIEFARSEAFQK	285
QY	121	SHGRGPEARLLVVVTGDESHDGEPLPAALKACACAGRVTTRYGVIAVLGHVLRQDRPSSFL	180
DB	286	--GGRKGAKKVMIVITDGESHSDPLEKVIQOOSRDNVTRYAVAVLYNRRGINPETFL	343
QY	181	REIRTIASDPDRFFNV	198
DB	344	NEIKIASDPDKHFFNV	361

RESULT 2

US-09-000-004A-2
; Sequence 2, Application US/09000004A
; Patent No. 6780603

```
; GENERAL INFORMATION:
; APPLICANT: Tsalibary, Photini-Effie
; APPLICANT: Charonis, Aristidis S.
; APPLICANT: Setty, Suman
; APPLICANT: Mauel, Michael
; TITLE OF INVENTION: ANALYSIS OF ALPHA INTEGRINS FOR THE DIAGNOSIS OF DIABETIC NEPHRO
; FILE REFERENCE: 600.314USWO
; CURRENT APPLICATION NUMBER: US/09/000.004A
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/001,387
; PRIOR FILING DATE: 1995-07-21
; PRIOR APPLICATION NUMBER: US 60/001,861
; PRIOR FILING DATE: 1995-08-03
; PRIOR APPLICATION NUMBER: US 60/016,700
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: PCT/US96/12067
; PRIOR FILING DATE: 1996-07-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Rattus
US-09-000-004A-2

Query Match 45.4%; Score 465.5; DB 4; Length 1180;
Best Local Similarity 46.5%; Pred. No. 3.3e-49;
Matches 92; Conservative 35; Mismatches 70; Indels 1; Gaps 1;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLDIP 60
Db 146 GICSDVPTFQVNSFAP-VQECSTQLDIVLDGNSIYPWSEVIAFLNDLLKRMMDIGP 204
Qy 61 EQIOVGIVOGESPVHWSLGDFTKEEVVRAAKNLSRREGRETAKTAQAIWVACTEGFSQ 120
Db 205 KQTQGVIVQGENVTHFNLNKYSSTEEVLVAANKIGRQGLQTMALGIDTARKEAFTE 264
Qy 121 SHGGRPEAARLLVVVTGDESHDGBELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 265 ARGARRGVKKVMVITDGHSHDNYLKVQIQDCEDENIQRFSAIILGHYNRGNLSTEFV 324
Qy 181 RERTIASDPDERFFNV 198
Db 325 EEIKSIASEPTEKHFNV 342

RESULT 3
US-10-061-658-5
; Sequence 5, Application US/10061658
; Patent No. 6652856
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US
; CURRENT APPLICATION NUMBER: US/10/061,658
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Rat
US-10-061-658-5

Query Match 44.1%; Score 452.5; DB 4; Length 214;
Best Local Similarity 46.6%; Pred. No. 1e-48;
Matches 90; Conservative 34; Mismatches 68; Indels 1; Gaps 1;

; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US
```

```
Qy 6 VDASFQPGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLDIP 65
Db 1 VSPTFQVNSFAP-VQECSTQLDIVLDGNSIYPWSEVIAFLNDLLKRMMDIGPKQTQV 59
Qy 66 GLVOYGESPVHWSLGDFTKEEVVRAAKNLSRREGRETAKTAQAIWVACTEGFSQSHGGR 125
Db 60 GIVOYGENVTHFNLNKYSSTEEVLVAANKIGRQGLQTMALGIDTARKEAFTEARGAR 119
Qy 126 PEARLLVVVTGDESHDGBELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFLREIRT 185
Db 120 RGVKKVMVITDGHSHDNYLKVQIQDCEDENIQRFSAIILGHYNRGNLSTEFVVEIKS 179
Qy 186 IASDPDERFFNV 198
Db 180 IASEPTEKHFNV 192

RESULT 4
US-10-061-658-6
; Sequence 6, Application US/10061658
; Patent No. 6652856
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US
; CURRENT APPLICATION NUMBER: US/10/061,658
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-061-658-6

Query Match 43.9%; Score 449.5; DB 4; Length 214;
Best Local Similarity 46.1%; Pred. No. 2.4e-48;
Matches 89; Conservative 35; Mismatches 68; Indels 1; Gaps 1;

Qy 6 VDASFQPGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLDIP 65
Db 1 VSPTFQVNSIAP-VQECSTQLDIVLDGNSIYPWSDSVTAFLNDLLKRMMDIGPKQTQV 59
Qy 66 GLVOYGESPVHWSLGDFTKEEVVRAAKNLSRREGRETAKTAQAIWVACTEGFSQSHGGR 125
Db 60 GIVOYGENVTHFNLNKYSSTEEVLVAANKIVQGRQTMALGIDTARKEAFTEARGAR 119
Qy 126 PEARLLVVVTGDESHDGBELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFLREIRT 185
Db 120 RGVKKVMVITDGHSHDNYLKVQIQDCEDENIQRFSAIILGHYNRGNLSTEFVVEIKS 179
Qy 186 IASDPDERFFNV 198
Db 180 IASEPTEKHFNV 192

RESULT 5
US-10-061-658-9
; Sequence 9, Application US/10061658
; Patent No. 6652856
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliarsky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US
```

; CURRENT APPLICATION NUMBER: US/10/061,658
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-061-658-9

Query Match 43.9%; Score 449.5; DB 4; Length 214;
Best Local Similarity 46.1%; Pred. No. 2.4e-48;
Matches 89; Conservative 35; Mismatches 68; Indels 1; Gaps 1;
Qy 6 VDASFQPGSLAPTAQRCPTMVDVVLVLDGNSIYPWSEVOTFLRLVGLKLFIDPEQIQV 65
Db 1 VSPFTQVNSIAP-VQSCSTQLDIVVLVLDGNSIYPWSDVTAFLNDLLKRMIDIGPKQTQV 59
Qy 66 GLVGYGSPVHESLGPRTKEEVVRAAKNLSRREGRETKTAQAIMVACTGFGSQSHGGR 125
Db 60 GIVGYGENVTHEFLNLYKSYSTEELVAAKTIQVGRGRTWTALGTDTRKEAFTARGAR 119
Qy 126 PEARLLVVVTDGSHGDEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFLREIRT 185
Db 120 RGVKKVMVIVTDGSHDNHRLKKVIQDCDENIQRFSAIILGNSYRGNLSKEFVEBIKS 179
Qy 186 IASDPDERFFNNV 198
Db 180 IASEPTEKHFFNV 192

RESULT 6
US-09-000-004A-4
; Sequence 4, Application US/090000004A
; Patent No. 6780603
; GENERAL INFORMATION:
; APPLICANT: Tslibary, Photini-Effie
; APPLICANT: Charonia, Aristidis S.
; APPLICANT: Setty, Suman
; APPLICANT: Mauer, Michael
; TITLE OF INVENTION: ANALYSIS OF ALPHA INTEGRINS FOR THE DIAGNOSIS OF DIABETIC NEPHRO
; FILE REFERENCE: 600.314USWO
; CURRENT APPLICATION NUMBER: US/09/000,004A
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/001,387
; PRIOR FILING DATE: 1995-07-21
; PRIOR APPLICATION NUMBER: US 60/001,861
; PRIOR FILING DATE: 1995-08-03
; PRIOR APPLICATION NUMBER: US 60/016,700
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: PCT/US96/12067
; PRIOR FILING DATE: 1996-07-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-000-004A-4

Query Match 43.0%; Score 441; DB 4; Length 1181;
Best Local Similarity 44.4%; Pred. No. 4.2e-46;
Matches 88; Conservative 33; Mismatches 77; Indels 0; Gaps 0;
Qy 1 GICARVDASFQPGSLAPTAQRCPTMVDVVLVLDGNSIYPWSEVOTFLRLVGLKLFIDP 60
Db 147 GVCSDISPDFQLSASFSPATQPCPSLDVVVCDNSIYPWDVAKNLFKVFQGLDIGP 206
Qy 61 EQIQVGLVQGESPVHESLGDFTKEEVVRAAKNLSRREGRETKTAQAIMVACTGFSQ 120

Db 207 TKTVQGLIQYANPRVFNLTNTYKTEEMIVATSTQSYGGDLTNTGAIQYARKYAYSA 266
Qy 121 SHGGRPEARLLVVVTDGSHGDEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 267 ASGGRRSATKVMVVVTDGSHGDEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 326
Qy 181 REIRTIASDPDERFFNNV 198
Db 327 KEIKAIASIPTRYFFNV 344

RESULT 7
US-09-949-016-6189
; Sequence 6189, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6189
; LENGTH: 1181
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6189

Query Match 43.0%; Score 441; DB 4; Length 1181;
Best Local Similarity 44.4%; Pred. No. 4.2e-46;
Matches 88; Conservative 33; Mismatches 77; Indels 0; Gaps 0;
Qy 1 GICARVDASFQPGSLAPTAQRCPTMVDVVLVLDGNSIYPWSEVOTFLRLVGLKLFIDP 60
Db 147 GVCSDISPDFQLSASFSPATQPCPSLDVVVCDNSIYPWDVAKNLFKVFQGLDIGP 206
Qy 61 EQIQVGLVQGESPVHESLGDFTKEEVVRAAKNLSRREGRETKTAQAIMVACTGFSQ 120
Db 207 TKTVQGLIQYANPRVFNLTNTYKTEEMIVATSTQSYGGDLTNTGAIQYARKYAYSA 266
Qy 121 SHGGRPEARLLVVVTDGSHGDEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 267 ASGGRRSATKVMVVVTDGSHGDEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 326
Qy 181 REIRTIASDPDERFFNNV 198
Db 327 KEIKAIASIPTRYFFNV 344

RESULT 8
US-09-949-016-10747
; Sequence 10747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10747
; LENGTH: 1195
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10747

Query Match      43.0%; Score 441; DB 4; Length 1195;
Best Local Similarity 44.4%; Pred. No. 4.3e-46;
Matches 88; Conservative 33; Mismatches 77; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVLVDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 161 GVCSDISPDQLSASFSAPATPCPSLIDVVVCDNSIYPPWDVKNFLEKFKVQGLDIGP 220

Qy 61 EQIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRREGRETKTAAQAIMVACTEGFSQ 120
Db 221 TKTQVGLIQYANNPRVFNLTYYKKEIMIVATSTQSYGGDLTNTFGAIVGLYLRNLDTKNLI 340

Qy 121 SHGCRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 281 ASGRRSATKVMVVVTDGSHDGSMLKAVIDQCNDHNLRFGLAVGLYLRNLDTKNLI 340

Qy 181 REIRTIASDPDERFFNV 198
Db 341 KEIKAIASIPTRYFFNV 358

RESULT 9
US-09-532-310B-6
; Sequence 6, Application US/09532310B
; Patent No. 6596276
; GENERAL INFORMATION:
; APPLICANT: Senger, Donald R
; Detmar, Michael
; Claffey, Kevin P
; TITLE OF INVENTION: Method for inhibiting tumor
; angiogenesis in a living subject
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, Esq.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: Dell PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Microsoft Word version 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/532,310B
; FILING DATE: 22-Mar-2000
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker, Esq.
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: BIS-036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-532-310B-6

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10747
; LENGTH: 1195
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10747

Query Match      41.0%; Score 420.5; DB 4; Length 1183;
Best Local Similarity 43.2%; Pred. No. 1.7e-43;
Matches 86; Conservative 33; Mismatches 79; Indels 1; Gaps 1;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVLVDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 147 GVCSDISPDQLSASFSAPATPCPSLIDVVVCDNSIYPPWDVKNFLEKFKVQGLDIGP 206

Qy 61 EQIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRREGRETKTAAQAIMVACTEGFSQ 120
Db 207 TKTQVGLIQYANNPRVFNLTYYKKEIMIVATSTQSYGGDLTNTFGAIVGLYLRNLDTKNLI 326

Qy 121 SHGCRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 179
Db 267 ASGRRSATKVMVVVTDGSHDGSMLKAVIDQCNDHNLRFGLAVGLYLRNLDTKNLI 326

Qy 180 REIRTIASDPDERFFNV 198
Db 327 KEIKAIASIPTRYFFNV 345

RESULT 10
US-09-532-310B-5
; Sequence 5, Application US/09532310B
; Patent No. 6596276
; GENERAL INFORMATION:
; APPLICANT: Senger, Donald R
; Detmar, Michael
; Claffey, Kevin P
; TITLE OF INVENTION: Method for inhibiting tumor
; angiogenesis in a living subject
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, Esq.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: Dell PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Microsoft Word version 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/532,310B
; FILING DATE: 22-Mar-2000
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker, Esq.
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: BIS-036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-532-310B-5

Query Match      38.6%; Score 396; DB 4; Length 1183;
Best Local Similarity 42.9%; Pred. No. 2.1e-40;
Matches 85; Conservative 33; Mismatches 78; Indels 2; Gaps 2;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVLVDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 148 GICSDVSPTFQVNVNAP-VQECSTXLDIVLVDGNSIYPPMDSV-TALNDLLKRMIDGP 205
```

```

Qy 61 EQIOVGLVOYGESPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKTAQAIMVACTEGFSQ 120
Db 206 KXTXGVIVXYGENTHFNLNKYSSTEEVLVAAKKIVXRGGRXTWTALGTDTARKEAFTE 265
Qy 121 SHGRPEAARLLVVVTGESHGDBELPAALKACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 266 ARGARGVKVMVITDGEHDXNHRLLKKVICDCEDENIXRFSIALSGSYNRGNLSTKFFV 325
Qy 181 REIRTIASDDPDERFFNV 198
Db 326 EEIKSIASEPTKSFNV 343

RESULT 11
US-08-462-128-37
; Sequence 37, Application US/08462128
; Patent No. 5686059
; GENERAL INFORMATION:
; APPLICANT: Goetinck, Paul F.
; APPLICANT: Tondravi, Mehrdad
; TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,128
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/006,096
; FILING DATE: 15-JAN-1993
; APPLICATION NUMBER: US 07/866,403
; FILING DATE: 10-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-462-128-37
Query Match 20.0%; Score 205; DB 1; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

Qy 28 DWIVLDGNSIYP--WSEVQTFLLRLVKGKLFIDPEQIQVGLVOYGESPVHWSLGDFT 85
Db 275 DLVFLIDGSKSVRPENFELVKKFSIQIVDTLDVSDKLAQVGLVQYSSSVRQEPFLGRHT 334
Qy 86 KEEVVRAAKNLSRRREGRETAKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESHGDEE 145
Db 335 KKOIKAAVRNNSYME-KGTWMTGAALKYLIDNSFTVSSGARPGAKQKVGIVFTDGRSQDYIN 393
Qy 146 LPAALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLREIRTIASDP-DERFFF 196
Db 394 -DAAKKAKDLGFKMFAVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 12
US-08-463-180-37
; Sequence 37, Application US/08463180
; Patent No. 5741670
; GENERAL INFORMATION:
; APPLICANT: Goetinck, Paul F.
; APPLICANT: Tondravi, Mehrdad
; TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,180
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/006,096
; FILING DATE: 15-JAN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/866,403
; FILING DATE: 10-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-005DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-463-180-37
Query Match 20.0%; Score 205; DB 1; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

Qy 28 DWIVLDGNSIYP--WSEVQTFLLRLVKGKLFIDPEQIQVGLVOYGESPVHWSLGDFT 85
Db 275 DLVFLIDGSKSVRPENFELVKKFSIQIVDTLDVSDKLAQVGLVQYSSSVRQEPFLGRHT 334
Qy 86 KEEVVRAAKNLSRRREGRETAKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESHGDEE 145
Db 335 KKOIKAAVRNNSYME-KGTWMTGAALKYLIDNSFTVSSGARPGAKQKVGIVFTDGRSQDYIN 393
Qy 146 LPAALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLREIRTIASDP-DERFFF 196
Db 394 -DAAKKAKDLGFKMFAVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 13
US-08-001-078A-1
; Sequence 1, Application US/08001078A
; Patent No. 5872094
; GENERAL INFORMATION:
; APPLICANT: Goetinck, Paul F.
; APPLICANT: Tondravi, M., Mehrdad

```

```
; APPLICANT: Binette, Francois
; TITLE OF INVENTION: METHODS FOR PROMOTING CARTILAGE MATRIX
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, Suite 510
; CITY: BOSTON
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/001,078A
; FILING DATE: 06-JAN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-001-078A-1

Query Match 20.0%; Score 205; DB 2; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

Qy 28 DVVLVDGNSIYP--WSEVQTFRLRLVGLKFLIDPEIQVGLVQYGSPPVHWSLGDPRFT 85
Db 275 DLVFLIDGSKSVRPENFELVKFISQIVDTLDVSDKLAQVGLVQYSSVRQEPFLGRFHT 334

Qy 86 KEEVVRAAKNLSRRREGTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTDGESHGDEE 145
Db 335 KKDIAAARNMSYNE-KGTMTGAALKYILDNSFTVSSGARPGAKVGIVFTDGRSQDYIN 393

Qy 146 LPALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLREIRTIASDP-DERFFF 196
Db 394 -DAAKKADLGFKMFAVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 14
US-08-897-443-4
; Sequence 4, Application US/08897443
; Patent No. 5981263
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Kaser, Mathew
; TITLE OF INVENTION: HUMAN MATRILIN-3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,443
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0348 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1732121
; US-08-897-443-4

Query Match 20.0%; Score 205; DB 2; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

Qy 28 DVVLVDGNSIYP--WSEVQTFRLRLVGLKFLIDPEIQVGLVQYGSPPVHWSLGDPRFT 85
Db 275 DLVFLIDGSKSVRPENFELVKFISQIVDTLDVSDKLAQVGLVQYSSVRQEPFLGRFHT 334

Qy 86 KEEVVRAAKNLSRRREGTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTDGESHGDEE 145
Db 335 KKDIAAARNMSYNE-KGTMTGAALKYILDNSFTVSSGARPGAKVGIVFTDGRSQDYIN 393

Qy 146 LPALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLREIRTIASDP-DERFFF 196
Db 394 -DAAKKADLGFKMFAVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 15
US-08-463-218-1
; Sequence 1, Application US/08463218
; Patent No. 5986052
; GENERAL INFORMATION:
; APPLICANT: Goetinck, Paul F.
; APPLICANT: Tondravi, Mehrdad
; APPLICANT: Binette, Francois
; TITLE OF INVENTION: METHODS FOR PROMOTING CARTILAGE MATRIX
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, Suite 510
; CITY: BOSTON
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,218
; FILING DATE: 05-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/001,078
; FILING DATE: 06-JAN-1993
```

ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-008DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-463-218-1

Query Match 20.0%; Score 205; DB 2; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;
Qy 28 DWIVLDGNSIYP--WSEVQTEFLRLVGLKLFIDPEIQIQLVQYGESPVHEWSLGDFT 85
Db 275 DLVFLIDGSKSVRPENFELVKKFIQIVDTLDVSDKLAQVGLVQYSSVRQEFPLGRFHT 334
Qy 86 KEEVVRAAKNLRSREGRETAKTQAIMVACTEGFSQSHGGRPEAAARLLVVVTGDGSHDGE 145
Db 335 KKDIAKAAVRNNSYNE-KCTMTGAALKYLLIDNSFTVSSGARGPAQKVGIVFTDGRSQDYIN 393
Qy 146 LPAALKACEAG-RVTRYGIIVLGHYLRQRDPSPFLREIRTIASDP-DERFFF 196
Db 394 -DAAKAKADLGKMFVGVG-----NAVEDELREIASEPVAEHYFY 433

Search completed: April 6, 2005, 12:24:22
Job time : 9.45027 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 5.18011 Seconds
(without alignments)
3459.921 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEHPYCTLPVGPCEPKTTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trenbl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	192	100.0	1167	1	ITAG HUMAN
2	175	91.1	288	2	QB8M12
3	65.5	34.1	301	1	NODP_AZOB
4	59	30.7	823	2	Q8WY18
5	59	30.7	1189	1	ITAG HUMAN
6	58.5	30.5	2275	2	Q8QS41
7	58	30.2	444	2	Q8NZW4
8	58	30.2	496	2	Q7PTW2
9	58	30.2	497	2	Q7PTJ9
10	58	30.2	695	2	Q9YI02
11	57.5	29.9	3338	2	Q82PH6
12	56	29.2	94	2	Q9YTS1
13	56	29.2	283	2	Q842M3
14	56	29.2	404	2	Q6FSB8
15	56	29.2	600	1	ACEK_BORPA
16	56	29.2	823	2	Q8CE84
17	56	29.2	1188	1	ITAG_MOUSE
18	56	29.2	1188	2	Q7TQC3
19	55	28.6	321	2	Q6C493
20	55	28.6	892	2	Q7WXF8
21	55	28.6	1054	2	Q9SNY7
22	55	28.6	1940	2	Q6GPD0
23	54.5	28.4	299	1	NODP_RH1SB
24	54.5	28.4	302	2	Q8L3E5
25	54.5	28.4	302	2	Q8P6D0
26	54.5	28.4	302	2	Q8PHD0
27	54.5	28.4	403	2	Q7UW11
28	54.5	28.4	445	2	Q7XVK9
29	54.5	28.4	1173	2	Q8ZX19
30	54	28.1	215	2	Q8AGB0
31	54	28.1	217	2	Q8AGB1
					Q75578 homo sapien
					QB8m12 mus musculu
					P28603 azospirillu
					Q8WY18 homo sapien
					Q8WY18 homo sapien
					Q8QS41 pongine her
					Q8NZW4 brachydanio
					Q7PTM2 anopheles g
					Q7PTJ9 anopheles g
					Q9YI02 brachydanio
					Q82PH6 streptomyce
					Q9YTS1 human immun
					Q842M3 rhodococcus
					Q6FSB8 candida gla
					Q7W2P4 bordetella
					Q8CE84 mus musculu
					P61622 mus musculu
					Q7TQC3 mus musculu
					Q6C493 yarrowia li
					Q7WXF8 alcaligenes
					Q9SNY7 nicotiana t
					Q6GPD0 xenopus lae
					O07308 rhizobium s
					Q8L3E5 xanthomonas
					Q8P6D0 xanthomonas
					Q8PHD0 xanthomonas
					Q7UW11 rhodopirell
					Q7XVK9 oryza sativ
					Q8ZX19 pyrobaculum
					Q8AGB0 human immun
					Q8AGB1 human immun

RESULT 1

ID	ITAG HUMAN	STANDARD;	PRT;	1167 AA.
AC	Q75578; Q9UH28;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Integrin alpha-10 precursor.			
GN	Name=ITGA10;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Articular chondrocytes;			
RX	MEDLINE=88352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;			
RA	Camper L., Hellman U., Lundgren-Akerlund E.;			
RT	"Isolation, cloning, and sequence analysis of the integrin subunit alpha10, a beta1-associated collagen binding integrin expressed on chondrocytes.";			
RL	J. Biol. Chem. 273:20383-20389(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Endothelial cells, and Heart;			
RX	MEDLINE=20169197; PubMed=10702680;			
RA	Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,			
RA	Wang S.-X., Langley R., Krissansen G.W.;			
RT	"The integrin alpha10 subunit: expression pattern, partial gene structure, and chromosomal localization.";			
RT	Cytogenet. Cell Genet. 87:238-244(1999).			
CC	-!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.			
CC	-!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10 associates with beta-1.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: Widely expressed with highest expression in muscle and heart. Found in articular cartilage.			
CC	-!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins with I-domains do not undergo protease cleavage.			
CC	-!- SIMILARITY: Belongs to the integrin alpha chain family.			
CC	-!- SIMILARITY: Contains 7 FG-GAP repeats.			
CC	-!- SIMILARITY: Contains 1 VWFA domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; AF074015; AAC31952.1; -.			
CC	EMBL; AF112345; AAF21944.1; -.			
CC	EMBL; AF172723; AAF61638.1; -.			

ALIGNMENTS

32	54	28.1	682	2	Q8MMO3
33	54	28.1	699	2	Q15816
34	54	28.1	2148	1	VIT1_AEDAE
35	53.5	27.9	1041	2	Q72EY0
36	53	27.6	94	2	Q9YT21
37	53	27.6	408	2	Q88L14
38	53	27.6	714	2	Q9PU31
39	53	27.6	727	2	Q9YI03
40	53	27.6	1019	2	Q7XBC5
41	52	27.1	94	2	Q9YT23
42	52	27.1	94	2	Q9YT25
43	52	27.1	101	1	ATPK_YEAST
44	52	27.1	180	2	Q7Z462
45	52	27.1	308	2	Q82BP5

Q8mm03 dictyosteli
Q15816 dictyosteli
Q16927 aedes aegypt
Q72ey0 desulfovibr
Q9yt21 human immun
Q88l14 pseudomonas
Q9pu31 oreochromis
Q9yi03 brachydanio
Q7xbs5 viscum albu
Q9yt23 human immun
Q9yt25 human immun
Q06405 saccharomyc
Q7z462 homo sapien
Q82bp5 streptomyce

RT "Cloning, sequence analysis, and chromosomal localization of the novel
 human integrin alpha11 subunit (ITGA11).";
 Genomics 60:179-187(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal muscle, and Uterus;
 RX MEDLINE=93395147; PubMed=10464311; DOI=10.1074/jbc.274.36.25735;
 RA Velling T., Kusche-Gullberg M., Seijersen T., Gullberg D.;
 "cDNA Cloning and Chromosomal Localization of Human alpha(11)
 Integrin. A collagen-binding, i domain-containing, beta(1)-associated
 integrin alpha-chain present in muscle tissues.";
 J. Biol. Chem. 274:25735-25742(1999).
 [3]
 RN SEQUENCE OF 954-1188 FROM N.A.
 RP TISSUE=Fibroblast;
 RC Andreu N., Estivill X., Escarceller M., Sumoy L.;
 Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen.
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11
 CC associates with beta-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC heart, intermediate levels in skeletal muscle and intermediate to
 CC low levels in pancreas, kidney and placenta. According to Ref.1
 CC also found in brain, colon, lung, small intestine, stomach,
 CC testis, salivary glands, thyroid glands and prostate. Very low
 CC levels in peripheral blood lymphocytes, fetal brain and fetal
 CC liver.
 CC -1- DEVELOPMENTAL STAGE: Strongly up-regulated in differentiating
 CC fetal muscle cells (in vitro).
 CC -1- DOMAIN: The integrin I-domain (insert) is a VMFA domain. Integrins
 CC with I-domains do not undergo protease cleavage.
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -1- SIMILARITY: Contains 1 VMFA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF109681; AAF01258.1; -;
 DR EMBL; AF137378; AAD51919.2; -;
 DR EMBL; AL359064; CAB94392.1; -;
 DR HSP; F18614; IMHP.
 DR Genew; HGNC:6136; ITGA11.
 DR MIM; 604789; -;
 DR GO; GO:0008305; C:integrin complex; TAS.
 DR GO; GO:0005518; P:collagen binding; TAS.
 DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
 DR GO; GO:0007517; P:muscle development; TAS.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR02035; VMFA.
 DR Pfam; PF01839; FG-GAP; 3.
 DR Pfam; PF00092; VWFA; 1.
 DR PRINTS; PK01185; INTEGRINA.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWFA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
 DR PROSITE; PS50234; VWFA; 1.
 KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium;
 KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.
 FT SIGNAL 1 22 Potential.
 FT CHAIN 23 1189 Integrin alpha-11.
 FT DOMAIN 23 1142 Extracellular (Potential).
 FT TRANSMEM 1143 1165 Potential.
 FT DOMAIN 1166 1189 Cytoplasmic (Potential).
 FT REPEAT 38 94 FG-GAP 1.
 FT REPEAT 102 163 FG-GAP 2.

FT DOMAIN 164 345 VMFA.
 FT REPEAT 359 420 FG-GAP 3.
 FT REPEAT 422 475 FG-GAP 4.
 FT REPEAT 477 537 FG-GAP 5.
 FT REPEAT 539 598 FG-GAP 6.
 FT REPEAT 601 653 FG-GAP 7.
 FT DOMAIN 1154 1162 Poly-Leu.
 FT DOMAIN 1174 1177 Poly-Arg.
 FT CA_BIND 488 496 Potential.
 FT CA_BIND 551 559 Potential.
 FT CA_BIND 613 621 By similarity.
 FT DISULFID 76 83 By similarity.
 FT DISULFID 121 139 Potential.
 FT DISULFID 129 159 Potential.
 FT DISULFID 659 668 By similarity.
 FT DISULFID 674 729 By similarity.
 FT DISULFID 781 787 By similarity.
 FT DISULFID 881 893 By similarity.
 FT CARBOHYD 92 82 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 95 95 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 291 291 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 331 331 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 358 358 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 449 449 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 462 462 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 528 528 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 642 642 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 694 694 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 857 857 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 894 894 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 973 973 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 1032 1032 N-linked (GLCNAC. .) (Potential).
 FT CARBOHYD 1040 1040 N-linked (GLCNAC. .) (Potential).
 FT VARIANT 433 433 V -> M.
 FT VARIANT 524 524 /FTID=VAR_009889.
 FT VARIANT 972 972 R -> L.
 FT VARIANT 1003 1003 /FTID=VAR_009890.
 FT VARIANT 1030 1030 L -> P.
 FT VARIANT 1030 1030 /FTID=VAR_009891.
 FT VARIANT 1030 1030 I -> M.
 FT VARIANT 1030 1030 /FTID=VAR_009892.
 FT VARIANT 1094 1094 Missing.
 FT VARIANT 1094 1094 /FTID=VAR_009893.
 FT VARIANT 1094 1094 L -> V.
 FT SEQUENCE 1189 AA; 133609 MW; 60303C08A4A4CD52 CRC64;
 Query Match 30.7%; Score 59; DB 1; Length 1189;
 Best Local Similarity 41.7%; Pred. No. 41;
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 Qy 12 GPGPEFKTLRVONLGCYVVSGLI 35
 Db 981 GIGPFPSICFRIONLGLFPIHGIM 1004
 RESULT 6
 Q8QS41 PRELIMINARY; PRT; 2275 AA.
 ID Q8QS41
 AC Q8QS41;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Tegument protein UL48.
 OS Pongine herpesvirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=188763;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2421467; PubMed=12533697; DOI=10.1099/vir.0.18606-0;
 RA Davison A.J., Dolan A., Akter P., Addison C., Dargan D.J.,
 RA Alencor D.J., McGeoch D.J., Hayward G.S.;

RA Anopheles Genome Sequencing Consortium;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008799; EAA03815.2; -.
 DR HSSP; P14779; 1JPZ.
 DR GO; GO:0004497; P:monooxygenase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR InterPro; IPR002401; EP4501.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00463; EP4501.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 FT NON_TER 497 497
 FT NON_TER 1
 SQ SEQUENCE 497 AA; 57002 MW; CD558BFD2C9C701C CRC64;
 Query Match 30.2%; Score 58; DB 2; Length 497;
 Best Local Similarity 26.2%; Pred. No. 22;
 Matches 17; Conservative 4; Mismatches 8; Indels 36; Gaps 2;
 QY 2 EVHPYGLPVGPGP-----EPKTLRVQNL----- 26
 DB 427 ERHPYALPFGGPRNCIGIRYAWLSMKILLIAHLVRYRPTTLKMEDELKPAIILIRIM 486
 QY 27 -GCYV 30
 DB 487 NGCLV 491
 RESULT 10
 ID Q9YI02 PRELIMINARY; PRT; 695 AA.
 AC Q9YI02;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Nuclear oncoprotein.
 GN Name=skib; Synonyms=skib; (Danio rerio).
 OS Brachydanio rerio (Zebrafish)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20368164; PubMed=10906458; DOI=10.1016/S0925-4773(00)00351-8;
 RA Kaufman C.D., Martinez-Rodriguez G., Hackett P.B. Jr.;
 RT "Ectopic expression of c-ski disrupts gastrulation and neural
 RL patterning in zebrafish."
 RL Mech. Dev. 95:147-162(2000).
 DR EMBL; AF060118; AAC64707.1; -.
 DR HSSP; P12755; 1MR1.
 DR ZFIN; ZDB-GENE-990715-10; skib.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
 DR InterPro; IPR000961; Putativ DNA_bind.
 DR InterPro; IPR010919; SAND like.
 DR InterPro; IPR003380; Transform_ski.
 DR Pfam; PF02437; Ski_Sno1.
 SQ SEQUENCE 695 AA; 79135 MW; 9BD424C32F61C4ED CRC64;
 Query Match 30.2%; Score 58; DB 2; Length 695;
 Best Local Similarity 35.9%; Pred. No. 32;
 Matches 14; Conservative 4; Mismatches 11; Indels 10; Gaps 1;
 QY 5 PYGTLPGVGPGEF-----KTLRVQNLGCYVVG 33
 DB 58 PAQTFPVPMPGLFIPSRSTRCRTVLRERETISCFVVG 96

RESULT 11

Q82PH6 PRELIMINARY; PRT; 3338 AA.
 ID Q82PH6
 AC Q82PH6;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Putative transcriptional activator SRCAP homolog.
 GN OrderedLocNames=SAV926;
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680;
 RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
 RA Omura S., Ikeda H., Iehikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites."
 RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Iehikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis."
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005024; BAC68636.1; -.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 3338 AA; 345044 MW; 5254F9C38439054B CRC64;
 Query Match 29.9%; Score 57.5; DB 2; Length 3338;
 Best Local Similarity 56.5%; Pred. No. 2e+02;
 Matches 13; Conservative 2; Mismatches 7; Indels 1; Gaps 1;
 QY 8 TLPVGQPEPKTTLRVQNLGCYV 30
 DB 2091 TLP-GPDPEFSPALRLRLGLQV 2112
 RESULT 12
 Q9Y751 PRELIMINARY; PRT; 94 AA.
 ID Q9Y751
 AC Q9Y751;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN Name=env;
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98445411; PubMed=9770526; DOI=10.1073/pnas.95.21.12568;
 RA Markham R.B., Wang W.C., Weissstein A.E., Wang Z., Munoz A.,
 RA Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H.,
 RA Yu X.P.;
 RT "Patterns of HIV-1 evolution in individuals with differing rates of
 RT CD4 T cell decline."
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).
 DR EMBL; AF089156; AAC78866.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.

Genolevures:

Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C., Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Nicaud J.M., Nikoleki M., Ostas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Swennene D., Tekcia F., Wesolowski-Louvel M., Westhof E., Wirth B., Zentou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.L.;

"Genome evolution in yeasts.";

Nature 430:35-44(2004).

RL EMBL; CR380954; CAG59609.1; -

DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . ; IEA.

DR GO; GO:0006807; P:nitrogen metabolism; IEA.

DR InterPro; IPR003010; Ntlse/CNhydts.

DR Pfam; PF00795; CN hydrolase; 1.

DR PROSITE; PS0263; CN HYDROLASE; 1.

SQ SEQUENCE 404 AA; 45985 MW; 2C448299FB6130D6 CRC64;

Query Match 29.2%; Score 56; DB 2; Length 404;

Best Local Similarity 46.7%; Pred.No. 34;

Matches 14; Conservative 1; Mismatches 13; Indels 2; Gaps 2;

QY 5 PYGTLPGVGPBKFTLRVQNL-GCVVSG 33

DB 68 PY-VCKAGEGSPSELAKRVSKLFGCYTVIG 96

RESULT 15

ACEK_BORPA STANDARD; PRT; 600 AA.

ID ACEK_BORPA

AC Q7W2P4;

AT 29-MAR-2004 (Rel. 43, Created)

CT 29-MAR-2004 (Rel. 43, Last sequence update)

DT 25-OCT-2004 (Rel. 45, Last annotation update)

DE Isocitrate dehydrogenase kinase/phosphatase (EC 2.7.1.116)

DE (EC 3.1.3.-) (IDH kinase/phosphatase) (IDHK/P).

GN Name:acek; OrderedLocNames=BPP4360;

OS Bordetella parapertussis.

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Alcaligenaceae; Bordetella.

NCBI_TaxID=519;

LN [1]

SEQUENCE FROM N.A.

RC STRAIN=12822 / ATCC BAA-587;

RP MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;

RX Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R., Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L., Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A., Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I., Chillingworth T., Collins M., Cronin A., Davis P., Doggett J., Feltham T., Goble A., Hamlin N., Hauser H., Holroyd S., Jageis K., Leather S., Moulé S., Norberczak H., O'Neill S., Ormond D., Price C., Rabinovitch E., Rutter S., Sanders M., Saunders D., Seeger K., Sharp S., Simmonds M., Skellton J., Squares R., Squares S., Stevens K., Unwin L., Whitehead S., Barrall B.G., Maskell D.J.;

RT "Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica.";

RT Nat. Genet. 35:32-40(2003).

RL CC -!- FUNCTION: Bifunctional enzyme which can phosphorylate or dephosphorylate isocitrate dehydrogenase (IDH) on a specific serine residue. This is a regulatory mechanism which enables bacteria to bypass the krebs cycle via the glyoxylate shunt in response to the source of carbon. When bacteria are grown on glucose, IDH is fully active and unphosphorylated, but when grown on acetate or ethanol, the activity of IDH declines drastically concomitant with its phosphorylation (by similarity).

```

CC -|- CATALYTIC ACTIVITY: ATP + [isocitrate dehydrogenase (NADP+)] = ADP
CC + [isocitrate dehydrogenase (NADP+)] phosphate.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: Belongs to the aceK family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; BX640436; CAB39639.1; -.
CC DR HAMAP; MF_00747; -; 1.
CC DR InterPro; IPR010452; AceK.
CC DR Pfam; PF06315; AceK; 1.
CC KW ATP-binding; Complete proteome; Glyoxylate bypass; Hydrolase; Kinase;
CC KW Multifunctional enzyme; Protein phosphatase; Transferase;
CC KW Tricarboxylic acid cycle.
CC FT NP_BIND 335 341 ATP (By similarity).
CC FT BINDING 356 356 ATP (By similarity).
CC FT ACT_SITE 390 390 By similarity.
CC SQ SEQUENCE 600 AA; 68516 MW; A4F10F6987971226 CRC64;

Query Match 29.2%; Score 56; DB 1; Length 600;
Best Local Similarity 33.3%; Pred. No. 52;
Matches 13; Conservative 4; Mismatches 14; Indels 8; Gaps 1;

Qy 5 PYGTLVPVGPGEFKTTLRVQ-----NLGCYVVSGLI 35
Db 196 PYADLPDAGPRLASDCQIQVLGSLFFRNTGAYIVGRLI 234

```

Search completed: April 6, 2005, 12:15:25
Job time : 7.18011 secs

RESULT 3
S61172
hypothetical protein YDR377w - yeast (*Saccharomyces cerevisiae*)
N/Alternate names: hypothetical protein D9481.21
C/Species: *Saccharomyces cerevisiae*
C/Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C/Accession: S61172
R/Ding, H.
submitted to the EMBL Data Library, June 1995
A/Description: The sequence of *S. cerevisiae* cosmid 9481.
A/Reference number: S61159
A/Accession: S61172
A/Molecule type: DNA
A/Residues: 1-101 <DIN>
A/Cross-references: UNIPROT:Q06405; EMBL:U28373; PIDN:AAB64813.1; PID:Q8491
A/Experimental source: strain S288C (AB972)
C/Genetics:
A/Gene: SGD:ATP17
A/Cross-references: SGD:S0002785; MIPS:YDR377w
A/Map position: 4R
C/Superfamily: *Saccharomyces* hypothetical protein YDR377w

Query Match 27.1%; Score 52; DB 2; Length 101;
Best Local Similarity 47.6%; Pred. No. 4.1;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 3 VHPYGTLPVGPGEFKTTLRV 23
|||:|||||:
DB 34 VHFYKSLPQGPAPAKANTRL 54
|||:|||||:

RESULT 4
TVPVSK
transforming protein ski - avian erythroblastosis virus (strain Sloan-Kettering)
C/Species: avian erythroblastosis virus
C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jul-2004
C/Accession: A32574
R/Stavnezer, E.; Brodeur, D.; Brennan, L.A.
Mol. Cell. Biol. 9, 4038-4045, 1989
A/Title: The v-ski oncogene encodes a truncated set of c-ski coding exons with limited
A/Reference number: A32574; MUID:89384634; PMID:2674685
A/Accession: A32574
A/Molecule type: DNA
A/Residues: 1-437 <STA>
A/Cross-references: UNIPROT:P17863
C/Genetics:
A/Gene: v-ski
C/Superfamily: ski transforming protein
C/Keywords: transforming protein

Query Match 27.1%; Score 52; DB 1; Length 437;
Best Local Similarity 35.3%; Pred. No. 20;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVQNLGCVVWSG 33
|||:|||||:
DB 46 PVMPGPFMPDSRSTERCETILGETISCFVVG 79
|||:|||||:

RESULT 5
I53106
gene gli protein - mouse (fragment)
C/Species: Mus sp. (mouse)
C/Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 23-Feb-1997
C/Accession: I53106
R/Walterhouse, D.; Ahmed, M.; Slusarski, D.; Kalamaras, J.; Boucher, D.; Holmgren, R.; I
Dev. Dyn. 196, 91-102, 1993
A/Title: gli, a zinc finger transcription factor and oncogene, is expressed during norma
A/Reference number: I53106; MUID:93372381; PMID:8364225
A/Accession: I53106
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA

A/Residues: 1-564 <RES>
A/Cross-references: GB:S65038; NID:G410635; PID:G410636
C/Genetics:
A/Gene: gli
C/Superfamily: gli transforming protein

Query Match 27.1%; Score 52; DB 2; Length 564;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 YEVHPYGTLPVGP 15
|||:|||||:
DB 472 YEARGPGSUPLGPGP 486
|||:|||||:

RESULT 6
C32575
C-ski protein FB27 - chicken
C/Species: *Gallus gallus* (chicken)
C/Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 18-Jun-1993
C/Accession: C32575
R/Sutcliffe, P.; Hughes, S.H.
Mol. Cell. Biol. 9, 4046-4051, 1989
A/Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski ge
A/Reference number: A32575; MUID:89384635; PMID:2779576
A/Accession: C32575
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-649 <SUT>
A/Cross-references: GB:M28517
C/Superfamily: ski transforming protein

Query Match 27.1%; Score 52; DB 2; Length 649;
Best Local Similarity 35.3%; Pred. No. 31;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVQNLGCVVWSG 33
|||:|||||:
DB 67 PVMPGPFMPDSRSTERCETILGETISCFVVG 100
|||:|||||:

RESULT 7
B32575
C-ski protein FB28 - chicken
C/Species: *Gallus gallus* (chicken)
C/Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 18-Jun-1993
C/Accession: B32575
R/Sutcliffe, P.; Hughes, S.H.
Mol. Cell. Biol. 9, 4046-4051, 1989
A/Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski ge
A/Reference number: A32575; MUID:89384635; PMID:2779576
A/Accession: B32575
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-713 <SUT>
A/Cross-references: GB:M28517
C/Superfamily: ski transforming protein

Query Match 27.1%; Score 52; DB 2; Length 713;
Best Local Similarity 35.3%; Pred. No. 35;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVQNLGCVVWSG 33
|||:|||||:
DB 67 PVMPGPFMPDSRSTERCETILGETISCFVVG 100
|||:|||||:

RESULT 8
TVHUSK
transforming protein ski - human
C/Species: *Homo sapiens* (man)
C/Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 09-Jul-2004
C/Accession: S06053

A48753

NFAT transcription factor phosphoprotein - mouse (fragment)
N;Alternate names: nuclear factor of activated T cells, phosphoprotein chain
C;Species: Mus musculus (house mouse)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999
C;Accession: A48753
R;McCaifrey, P.G.; Luo, C.; Kerppola, T.K.; Jain, J.; Badalian, T.M.; Ho, A.M.; Burgeon, S.; Science 262, 750-754, 1993
A;Title: Isolation of the cyclosporin-sensitive T cell transcription factor NFATp.
A;Reference number: A48753; MUID:94053710; PMID:8235597
A;Accession: A48753
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-890 <MCC>
A;Cross-references: GB:U02079
C;Keywords: phosphoprotein; transcription factor

Query Match 26.6%; Score 51; DB 2; Length 890;
Best Local Similarity 75.0%; Pred. No. 61;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 HPVGTLPVGPQP 15

DB 805 HPLGTCPLPQP 816

RESULT 14

S34172
sucrose-phosphate synthase (EC 2.4.1.14) - potato
C;Species: Solanum tuberosum (potato)
C;Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 09-Jul-2004
C;Accession: S34172
R;Sonnewald, U.; Basner, A.
submitted to the EMBL Data Library, June 1993
A;Reference number: S34172
A;Accession: S34172
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1053 <SON>
A;Cross-references: UNIPROT:Q43845; EMBL:X73477; NID:G313264; PIDN:CAA51872.1; PID:G313264
C;Function:
A;Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fructose
A;Pathway: sucrose biosynthesis
C;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C;Keywords: Glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F;168-651/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 26.6%; Score 51; DB 2; Length 1053;
Best Local Similarity 35.5%; Pred. No. 74;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEHPYGTLPVGPPEPKTTLRVONLGYVY 31

DB 911 FKVKPGTVP--PSKELKVMRIQALRCHAV 939

RESULT 15

T03464
probable methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) / / methenyltetrahydrofolate dehydrogenase
C;Species: Rhodospirillum rubrum
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C;Accession: T03464
R;Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A;Title: Sequence of a 189-kb segment of the chromosome of Rhodospirillum rubrum
A;Reference number: 214955; MUID:97404404; PMID:9256491
A;Accession: T03464
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA
A;Residues: 1-300 <VLC>
A;Cross-references: UNIPROT:O68031; EMBL:AF010496; NID:G3128256; PIDN:AAC16117.1; PID:G3128256
C;Genetics:
A;Map position: 1

C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate dehydrogenase; hydrolase; multifunctional enzyme; oxidoreductase
C;Keywords: hydrolase; multifunctional enzyme; oxidoreductase
F;9-287/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>
Query Match 26.3%; Score 50.5; DB 2; Length 300;
Best Local Similarity 38.5%; Pred. No. 22;
Matches 10; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

QY 2 EVHPYGTLPVGPPEPKTTLRVQ 24

DB 32 EHPVGLVLSISIGSPSEIEVYVRNQ 57

Search completed: April 6, 2005, 12:17:01
Job time : 3.21966 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 4.42639 Seconds
(without alignments)
3058.161 Million cell updates/sec

Title: us-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEVHPYGTLPVGPGEKTTLRVQNLGCYVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	100.0	1049	5	ADR41424 Human CD-
2	192	100.0	1152	4	AAB64657 Human sec
3	192	100.0	1152	4	AAB64658 Human sec
4	192	100.0	1167	3	AY32242 Human int
5	192	100.0	1167	4	AAB64584 Human sec
6	192	100.0	1167	6	ABP99490 Human sec
7	192	100.0	1167	6	ABR00964 Human gen
8	192	100.0	1167	6	ADA44026 Human sec
9	192	100.0	1167	8	ADQ19290 Human sof
10	142	74.0	1132	3	AY32243 Human int
11	69	35.9	1188	4	AAU14467 Human nov
12	69	35.9	1188	4	AAU14231 Human nov
13	69	35.9	1188	4	AAB50085 Human A25
14	69	35.9	1188	5	AAU10551 Human A25
15	69	35.9	1188	7	ADE09956 Novel pro
16	69	35.9	1189	8	ADH80785 Human pol
17	59	30.7	1034	3	AAB25590 Protein e
18	59	30.7	1034	6	ADA27062 Human nov
19	59	30.7	1034	8	ADE86592 Novel hum
20	59	30.7	1058	5	ADR41496 Human CD-
21	59	30.7	1120	6	ABR58365 Human NOV
22	59	30.7	1188	4	AAB30929 Amino aci
23	59	30.7	1189	3	AAB25582 ITG11 pr
24	59	30.7	1189	4	ABG12949 Novel hum
25	59	30.7	1189	6	ABR58364 Human NOV

26	59	30.7	1189	6	ADA27054	AdA27054 Human nov
27	59	30.7	1189	7	ADE63570	Ade63570 Human Pro
28	59	30.7	1189	8	ADE86584	Ade86584 Novel hum
29	59	30.7	1189	8	ADQ19968	Adq19968 Human sof
30	56	29.2	266	7	ABM85385	ABm85385 Human pro
31	56	29.2	347	2	AAU11069	Aau11069 12D3 anti
32	56	29.2	545	5	ABB72288	Abb72288 Murine pr
33	56	29.2	688	5	ABB72300	Abb72300 Rat prote
34	56	29.2	696	5	ABB72289	Abb72289 Rat prote
35	56	29.2	1188	4	AAB50087	Aab50087 Murine A2
36	56	29.2	1188	5	AAU10552	Aau10552 Murine A2
37	55.5	28.9	987	8	ADP29492	Adp29492 Human sec
38	55	28.6	210	4	AAU62597	Aau62597 Propionib
39	55	28.6	210	6	ABM59116	ABm59116 Propionib
40	54	28.1	257	4	ABG20619	ABg20619 Novel hum
41	54	28.1	391	4	ABG20618	ABg20618 Novel hum
42	54	28.1	2139	8	ADO17150	Ado17150 Novel exp
43	53.5	27.9	53	6	ABM64899	ABm64899 Propionib
44	53.5	27.9	349	4	ABG14393	ABg14393 Novel hum
45	53	27.6	408	6	ABU39798	ABu39798 Protein e

ALIGNMENTS

RESULT 1

ADR41424

ID ADR41424 standard; protein; 1049 AA.

XX ADR41424;

XX

DT 07-OCT-2004 (first entry)

XX

DE Human CD-like molecule HAOAD02, SEQ ID NO:223.

XX

Human; CD-like molecule; cluster of differentiation; diagnosis; prevention; immune disorder; immunodeficiency; autoimmune disorder; blood-related disorder; haematological disorder; haemostatic disorder; thrombolytic disorder; hyperproliferative disorder; cancer; tumour; apoptotic disorder; cardiovascular disorder; respiratory disorder; oncogenic disorder; neovascularisation; neurological disorder; endocrine disorder; reproductive system disorder; infectious disease; gastrointestinal disorder; drug screening; tissue regeneration; chemotaxis; gene therapy; antibody therapy; drug targeting; chromosome mapping; forensic analysis; immunophenotyping; cytostatic; haemostatic; tranquiliser; vulnery; antiinflammatory; nephrotropic; cardiant; antiallergic; anti-HIV; antirheumatic; antiarthritic; antipsoriatic; immunosuppressive; vasotropic; nootropic; neuroprotective; antithyroid; thyromimetic; gynaecological; virucide; hepatotropic; antibacterial; dermatological; chromosome lq21.

XX Homo sapiens.

OS WO200226930-A2.

PN

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029838.

XX

PR 26-SEP-2000; 2000US-0235484P.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Birse CE;

XX WPI; 2002-405050/43.

DR N-PSDB; ADR41248.

XX

Novel polynucleotides and polypeptides useful for treating, preventing or ameliorating cardiovascular, renal, neurovascular, and autoimmune disorders.

Claim 11; SEQ ID NO 223; 1243pp; English.

XX CC The invention relates to 167 novel human CD (cluster of differentiation)-
 CC like molecules (ADR41388-ADR41563) and to cDNAs encoding them (segid:11)-
 XX
 SQ Sequence 1049 AA;
 Query Match 100.0%; Score 192; DB 5; Length 1049;
 Best Local Similarity 100.0%; Pred. No. 1.2e-17;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YEHPYGTLPVGPPEFKTTLRVQNLGCVVVSGLI 35
 DB 834 YEHPYGTLPVGPPEFKTTLRVQNLGCVVVSGLI 868
 RESULT 2
 AAB64657
 ID AAB64657 standard; protein; 1152 AA.
 XX
 AC AAB64657;
 DT 22-MAR-2001 (first entry)
 XX
 DE Human secreted protein BLAST search protein SEQ ID NO: 167.
 XX
 OS Homo sapiens.
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO2000077197-A1.
 XX
 PD 21-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US014934.
 XX
 PR 11-JUN-1999; 99US-0138599P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis GA;
 XX
 DR WPI; 2001-032312/04.
 XX
 XX Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 XX
 PS Disclosure; Page 543-546; 558pp; English.
 XX
 CC The invention relates to the isolation of genes AAF32757-F32803 encoding
 CC the human secreted proteins AAB64549-B64594. The sequence is a search
 CC result from a BLASTX homology search. The genes and proteins are useful
 CC for preventing, ameliorating or treating medical conditions, e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment and
 CC prevention of: (a) cancer, e.g. breast and ovarian cancer, and other
 CC cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal
 CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
 CC disease, allergies, autoimmune haemolytic anaemia, autoimmune
 CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
 CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders
 CC such as myocardial ischaemias; (d) wound healing; (e) neurological
 CC diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases
 CC such as viral, bacterial, fungal and parasitic infections
 XX
 SQ Sequence 1152 AA;
 Query Match 100.0%; Score 192; DB 4; Length 1152;
 Best Local Similarity 100.0%; Pred. No. 1.2e-17;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 1.3e-17;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YEHPYGTLPVGPPEFKTTLRVQNLGCVVVSGLI 35
 DB 952 YEHPYGTLPVGPPEFKTTLRVQNLGCVVVSGLI 986
 RESULT 3
 AAB64658
 ID AAB64658 standard; protein; 1152 AA.
 XX
 AC AAB64658;
 XX
 DT 22-MAR-2001 (first entry)
 XX
 DE Human secreted protein BLAST search protein SEQ ID NO: 168.
 XX
 OS Homo sapiens.
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO2000077197-A1.
 XX
 PD 21-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US014934.
 XX
 PR 11-JUN-1999; 99US-0138599P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis GA;
 XX
 DR WPI; 2001-032312/04.
 XX
 XX Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 XX
 PS Disclosure; Page 547-551; 558pp; English.
 XX
 CC The invention relates to the isolation of genes AAF32757-F32803 encoding
 CC the human secreted proteins AAB64549-B64594. The sequence is used as a
 CC query sequence for doing BLASTX searches to identify homologous
 CC sequences. The genes and proteins are useful for preventing, ameliorating
 CC or treating medical conditions, e.g. by protein or gene therapy. The
 CC genes are isolated from a range of human tissues disclosed in the
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
 CC e.g. breast and ovarian cancer, and other cancers of the adrenal gland,
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections
 XX
 SQ Sequence 1152 AA;
 Query Match 100.0%; Score 192; DB 4; Length 1152;
 Best Local Similarity 100.0%; Pred. No. 1.3e-17;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YEHPYGTLPVGPPEFKTTLRVQNLGCVVVSGLI 35
 DB 952 YEHPYGTLPVGPPEFKTTLRVQNLGCVVVSGLI 986

RESULT 4
 AAY32242
 ID AAY32242 standard; protein; 1167 AA.
 XX
 AC AAY32242;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Human integrin subunit alpha-10.
 XX
 KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;
 KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..22
 FT /note= "signal peptide"
 FT Protein 23..1145
 FT /note= "mature protein"
 FT Domain 23..1120
 FT /note= "extracellular domain"
 FT Modified-site 98
 FT /note= "N-glycosylated"
 FT Domain 162..359
 FT /note= "I-domain"
 FT Modified-site 336
 FT /note= "N-glycosylated"
 FT Modified-site 364
 FT /note= "N-glycosylated"
 FT Binding-site 494..502
 FT /note= "cation binding site motif"
 FT Binding-site 558..566
 FT /note= "cation binding site motif"
 FT Binding-site 620..628
 FT /note= "cation binding site motif"
 FT Modified-site 733
 FT /note= "N-glycosylated"
 FT Modified-site 839
 FT /note= "N-glycosylated"
 FT Modified-site 921
 FT /note= "N-glycosylated"
 FT Modified-site 1018
 FT /note= "N-glycosylated"
 FT Modified-site 1039
 FT /note= "N-glycosylated"
 FT Domain 1121..1145
 FT /note= "transmembrane domain"
 FT Domain 1122..1167
 FT /note= "cytoplasmic domain, specifically claimed in Claim 21"
 XX
 PN WO9951639-A1.
 XX
 XX 14-OCT-1999.
 XX
 XX 31-MAR-1999; 99WO-SE000544.
 XX
 PR 02-APR-1998; 98SE-00001164.
 PR 28-JAN-1999; 99SE-00000319.
 XX
 XX (ACTI-) ACTIVE BIOTECH AB.
 XX
 XX Lundgren-Akerlund E;
 XX
 XX WPI; 2000-052639/04.
 DR N-PSDB; AAZ34719.
 XX
 XX New isolated integrin subunit alpha-10, used as a marker or target
 FT molecule for cells during development, regeneration and pathological

PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 XX inflammation.
 PS Claim 1; Fig 6; 90pp; English.
 XX
 CC This sequence represents novel human chondrocyte integrin subunit alpha-
 CC 10 (ISa10). A splice variant is given in AAY32243. The invention relates
 CC to a recombinant or isolated integrin heterodimer comprising the alpha10
 CC subunit in association with subunit beta (especially beta-1). The
 CC heterodimer and the subunit alpha-10 can be used as markers or targets of
 CC all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
 CC They can also be used: for treating pathological conditions involving
 CC ISa10, such as damage to cartilage, trauma, rheumatoid arthritis or
 CC osteoarthritis; for detecting the formation of cartilage during embryonal
 CC development, physiological or therapeutic repair of cartilage, or
 CC detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and analysis
 CC of for sorting, isolating or purification of chondrocytes; and as a target for
 CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or
 CC other tissues where adhesion impairs the function of the tissue (all
 CC claimed). ISa10 binding entities can be used to determine the
 CC differentiation-state of cells during embryonic development, angiogenesis
 CC or development of cancer, in pathological conditions such as rheumatoid
 CC arthritis, osteoarthritis or cancer, in tissue regeneration or in
 CC therapeutic and physiological repair of cartilage (claimed). A
 CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also
 CC claimed. ISa10 polynucleotides, vectors, host cells and methods of
 CC producing recombinant ISa10 are also claimed
 XX
 SQ Sequence 1167 AA;
 Query Match 100.0%; Score 192; DB 3; Length 1167;
 Best Local Similarity 100.0%; Pred. NO. 1.4e-17;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 YEYHPYCTLPVGPPEFKTTLRVQNLGCVVSGLI 35
 Db 952 YEYHPYCTLPVGPPEFKTTLRVQNLGCVVSGLI 986
 RESULT 5
 AAB64584
 ID AAB64584 standard; protein; 1167 AA.
 XX
 AC AAB64584;
 XX
 DT 22-MAR-2001 (first entry)
 XX
 XX Human secreted protein #37.
 DE
 DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 OS
 PN WO200077197-A1.
 XX
 XX 21-DEC-2000.
 PD
 XX 01-JUN-2000; 2000WO-US014934.
 PF
 XX 11-JUN-1999; 99US-0138599P.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX
 XX Rosen CA, Ruben SM, Komatsoulis GA;
 PI
 XX WPI; 2001-032312/04.
 DR

```
DR N-PSDB; AAF32793.
```

```
XX Isolated nucleic acid molecule encoding a human secreted protein is used
```

```
PT in preventing, treating or ameliorating a medical condition.
```

```
XX Claim 11; Page 496-500; 558pp; English.
```

```
XX Sequences AB64549-B64594 represent the amino acid sequences of 47 human
```

```
CC secreted proteins encoded by the genes AAF32757-F32803. The genes and
```

```
CC proteins are useful for preventing, ameliorating or treating medical
```

```
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
```

```
CC a range of human tissues disclosed in the specification. The nuclei
```

```
CC acids, proteins, antibodies and (ant)agonists are useful in the
```

```
CC diagnosis, treatment and prevention of:
```

```
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
```

```
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
```

```
CC immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic
```

```
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
```

```
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
```

```
CC cardiovascular disorders such as myocardial ischaemias;
```

```
CC / (e) neurological diseases such as cerebral anoxia and epilepsy; and (f)
```

```
CC infectious diseases such as viral, bacterial, fungal and parasitic
```

```
XX infections
```

```
SQ Sequence 1167 AA;
```

```
Query Match          100.0%; Score 192; DB 4; Length 1167;
```

```
Best Local Similarity 100.0%; Pred. No. 1.4e-17;
```

```
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 YEYHPYGTLPGPGEFKTLLRVQLNGCYVVSGLI 35  
   |||||  
Db 952 YEYHPYGTLPGPGEFKTLLRVQLNGCYVVSGLI 986  
   |||||
```

New human secreted proteins encoded by genes contained in cDNA clones (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS, multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or West Nile fever.

Claim 1; Page 1395-1398; 2423pp; English.

The invention relates to novel human genes (ABZ66891-ABZ68209) and the encoded secreted proteins (ABP99470-ABP99872) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections

Sequence 1167 AA;

Query Match 100.0%; Score 192; DB 6; Length 1167;
Best Local Similarity 100.0%; Pred. No. 1.4e-17;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEVHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 35
|||||
DB 952 YEVHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 986

RESULT 7

ABR00964
ID ABR00964 standard; protein; 1167 AA.
XX ABR00964;
XX
DT 12-MAY-2003 (first entry)
XX
DE Human gene 18-encoded secreted protein HAOAG15, SEQ ID NO:445.
XX
KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW autoimmune disorder; inflammation; angiogenic diseases; AIDS; HIV;
KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW drug screening; chromosome identification; chromosome mapping;
KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KW antianaemic; vulnarary; chromosome 1q21.
XX
OS Homo sapiens.
XX
XX WO200277013-A2.
XX
XX 03-OCT-2002.
XX
XX 26-MAR-2002; 2002WO-US009370.
XX
XX 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2003-040578/03.
DR
DR N-PSDB; ABZ73298.
XX
XX New human secreted proteins and nucleic acids, useful for detecting or treating cancer or other hyperproliferative disorders, autoimmune disorders, inflammatory disorders, HIV disease, hepatitis or anemia.

XX Claim 13; Page 1387-1390; 2474pp; English.

XX AB273281-AB273697 represent cDNAs corresponding to 391 human secreted

CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.

CC AB273698-AB274687 represent human secreted protein genomic fragments. The

CC invention also encompasses antibodies specific for the secreted proteins,

CC the use of the secreted proteins in drug screening and recombinant

CC vectors and host cells comprising a nucleic acid of the invention. The

CC secreted proteins are thought to be involved in biological activities

CC associated with cellular signalling, cellular differentiation, cell

CC migration, prohormone activation and neurotransmitter activity. The

CC secreted proteins, nucleic acids encoding them, antibodies or antibody

CC fragments specific for the secreted proteins, and modulators of protein

CC activity are useful for diagnosing or treating cancers or other

CC hyperproliferative disorders. Additionally, the secreted proteins and

CC their nucleic acids may also be used in the treatment of autoimmune

CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS

CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote

CC wound healing. Nucleic acids of the invention may be used for chromosome

CC identification, chromosome mapping, in gene therapy, for identifying

CC individuals from minute biological samples, as hybridisation probes, and

CC as molecular weight markers. The present sequence represents a human

CC secreted protein of the invention

XX

XX Sequence 1167 AA;

Query Match 100.0%; Score 192; DB 6; Length 1167;

Best Local Similarity 100.0%; Pred. No. 1.4e-17;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 35

Db 952 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 986

RESULT 8

ADA44026

ID ADA44026 standard; protein; 1167 AA.

XX AC ADA44026;

XX 20-NOV-2003 (first entry)

XX Human secreted protein SEQ ID 214.

DE Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;

XX Neuroprotective; Cerebroprotective; Antianemic.

XX Homo sapiens.

XX WO2003000865-A2.

XX 03-JAN-2003.

XX 26-MAR-2002; 2002WO-US009105.

XX 27-MAR-2001; 2001US-0278650P.

PR 12-SEP-2001; 2001US-00950082.

PR 12-SEP-2001; 2001US-00950083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-184045/18.

DR N-PSDB; ADA43832.

XX A human secreted protein and nucleic acids useful for preparing a

PT diagnostic or pharmaceutical composition for diagnosing or treating

PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,

PT retinopathy, neuropathy.

XX

PS Claim 1; SEQ ID NO 214; 701pp; English.

XX The invention relates to novel genes and their fragments which are useful

CC for preventing, treating or ameliorating medical conditions e.g. by

CC protein or gene therapy. The genes are isolated from a range of human

CC tissues disclosed in the specification. The nucleic acids and proteins

CC are useful in the diagnosis, treatment and prevention of conditions

CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,

CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,

CC infection, cataract, renal disorders, or endocrine disorders. The present

CC sequence was used to illustrate the invention.

XX

XX Sequence 1167 AA;

Query Match 100.0%; Score 192; DB 6; Length 1167;

Best Local Similarity 100.0%; Pred. No. 1.4e-17;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 35

Db 952 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 986

RESULT 9

ADQ19290

ID ADQ19290 standard; protein; 1167 AA.

XX AC ADQ19290;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.

DE soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

PR (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnik A;

XX WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression

PT of a gene in a first soft tissue sample and a normal soft tissue sample

PT and comparing the gene expression, also useful in treating soft tissue

PT sarcoma.

XX

XX Example 2; SEQ ID NO 2109; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual

CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level

CC of protein expression in the first soft tissue sample indicates the

CC presence of soft tissue sarcoma. The method of the invention has

CC cytostatic applications and may be useful for detecting soft tissue

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic

CC acid sequences may be useful in diagnostic and screening applications.

CC The current sequence is that of a human soft tissue sarcoma-upregulated

CC protein of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.

XX

XX Sequence 1167 AA;

Query Match 100.0%; Score 192; DB 8; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 1.4e-17;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEHPYGTLPVGPPEFKTTLRVQNLGCVVSGLI 35
 |||||
 DB 952 YEHPYGTLPVGPPEFKTTLRVQNLGCVVSGLI 986

RESULT 10
 AAY32243
 ID AAY32243 standard; protein; 1132 AA.
 XX
 AC AAY32243;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DX Human integrin subunit alpha-10 splice variant.
 XX
 KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;
 KW therapy; cartilage; chondrocytes; osteoblast; fibroblast; vaccine; marker;
 KW splice variant.
 XX
 OS Homo sapiens.
 XX
 FH Location/Qualifiers
 FT Peptide 1..22
 FT Protein /note= "signal peptide"
 FT /note= "mature protein"
 XX
 PN WO9951639-A1.
 XX
 PD 14-OCT-1999.
 XX
 XX 31-MAR-1999; 99WO-SE000544.
 PF
 PR 02-APR-1998; 98SE-00001164.
 PR 28-JAN-1999; 99SE-00000319.
 XX
 PA (ACTI-) ACTIVE BIOTECH AB.
 XX
 XX Lundgren-Akerlund E;
 PI
 XX
 DR WPI; 2000-052639/04.
 DR N-PSDB; AAZ34720.
 XX
 XX New isolated integrin subunit alpha-10, used as a marker or target
 PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation.
 XX
 XX Claim 1; Page 43-48; 90pp; English.
 PS
 XX
 XX This sequence represents a splice variant of novel human chondrocyte
 CC integrin subunit alpha-10 (ISa10). It is identical to ISa10 (see
 CC AAY32242) except for deletion of amino acids 975-986. The invention
 CC relates to a recombinant or isolated integrin heterodimer comprising the
 CC alpha10 subunit in association with subunit beta (especially beta-1). The
 CC heterodimer, subunit alpha-10 or splice variant can be used as a marker
 CC or target of all types of cells, e.g. of chondrocytes, osteoblasts and
 CC fibroblasts. They can also be used for treating pathological conditions
 CC involving ISa10, such as damage to cartilage, trauma, rheumatoid
 CC arthritis or osteoarthritis; for detecting the formation of cartilage
 CC during embryonal development, physiological or therapeutic repair of
 CC cartilage, or detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and analysis
 CC or for sorting, isolating or purification of chondrocytes and for in
 CC vitro studies of differentiation of chondrocytes; and as a target for
 CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or
 CC other tissues where adhesion impairs the function of the tissue (all

CC claimed). ISa10 binding entities can be used to determine the
 CC differentiation-state of cells during embryonic development, angiogenesis
 CC or development of cancer, in pathological conditions such as rheumatoid
 CC arthritis, osteoarthritis or cancer, in tissue regeneration or in
 CC therapeutic and physiological repair of cartilage (claimed). A
 CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also
 CC claimed. ISa10 polynucleotides, vectors, host cells and methods of
 CC producing recombinant ISa10 are also claimed

SQ Sequence 1132 AA;
 Query Match 74.0%; Score 142; DB 3; Length 1132;
 Best Local Similarity 83.3%; Pred. No. 1.2e-10;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 YEHPYGTLPVGPPEFKTTLRVQNLGCVV 30
 |||||
 DB 952 YEHPYGTLPVGPPEFKTTLRTNNASCIV 981

RESULT 11
 AAU14467
 ID AAU14467 standard; protein; 1188 AA.
 XX
 AC AAU14467;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DX Human novel protein #338.
 XX
 KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200155437-A2.
 XX
 PD 02-AUG-2001.
 XX
 XX 25-JAN-2001; 2001WO-US002623.
 PF
 XX 25-JAN-2000; 2000US-00491404.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX WPI; 2001-451939/48.
 DR
 DR N-PSDB; AAS22772.
 XX
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage.
 PT
 XX Example 4; Page 828-831; 894pp; English.
 PS
 XX The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.
 CC Polynucleotides of the invention are used as probes and primers, for
 CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise
 CC antibodies/ elicit an immune response, to determine quantitative protein

CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC represents a protein of the invention
 XX
 SQ Sequence 1188 AA;
 Query Match 35.9%; Score 69; DB 4; Length 1188;
 Best Local Similarity 40.5%; Pred. No. 1.9;
 Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;
 QY 1 YEVPYGTLP--VGPGEFKTTLRVQNLGCVVWSGLI 35
 DB 968 YEVPNSLERYDGIGPPFCIFRIQNLGLFPIHGM 1004
 RESULT 12
 AAU14231
 ID AAU14231 standard; protein; 1188 AA.
 AC AAU14231;
 DT 24-OCT-2001 (first entry)
 DE Human novel protein #102.
 KW Human; novel protein; Antianaemic; osteopathic; anti-inflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200155437-A2.
 XX
 XX 02-AUG-2001.
 XX
 XX 25-JAN-2001; 2001WO-US002623.
 XX
 XX 25-JAN-2000; 2000US-00491404.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 XX
 XX WPI; 2001-451939/48.
 XX
 XX N-PSDB; AAS22536.
 XX
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 XX nervous system disorders, and for regenerating bone and cartilage.
 XX
 XX Example 4; Page 578-581; 894pp; English.
 XX
 XX The invention relates to polynucleotides encoding novel human proteins or
 XX their active domains. The polypeptides, polynucleotides and antibodies
 XX raised against the polypeptides are used in a method of treatment of a
 XX mammal and prevention of disorders caused by the aberrant protein
 XX expression or activity. The polypeptides can be used as molecular weight
 XX markers, food supplements, and in antibody production. The polypeptides
 XX are used to identify compounds which bind to the polypeptides.
 XX Polynucleotides of the invention are used as probes and primers, for
 XX sequencing, for chromosome or gene mapping, in the production of

CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise
 CC antibodies/elicite an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC represents a protein of the invention
 XX
 SQ Sequence 1188 AA;
 Query Match 35.9%; Score 69; DB 4; Length 1188;
 Best Local Similarity 40.5%; Pred. No. 1.9;
 Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;
 QY 1 YEVPYGTLP--VGPGEFKTTLRVQNLGCVVWSGLI 35
 DB 968 YEVPNSLERYDGIGPPFCIFRIQNLGLFPIHGM 1004
 RESULT 13
 AAB50085
 ID AAB50085 standard; protein; 1188 AA.
 XX
 AC AAB50085;
 XX
 XX 19-MAR-2001 (first entry)
 DT Human A259.
 DE Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
 KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
 KW rheumatoid arthritis.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT Domain 1..1141
 FT /label= Extracellular_domain
 FT Peptide 1..22
 FT /label= Signal_peptide
 FT Protein 23..1188
 FT /label= Mature_protein
 FT Domain 39..74
 FT /label= Integrin_alphasubunit_repeat_domain_#1
 FT Domain 115..157
 FT /label= Integrin_alphasubunit_repeat_domain_#2
 FT Domain 164..345
 FT /label= I_domain
 FT Domain 367..392
 FT /label= Integrin_alphasubunit_repeat_domain_#3
 FT Domain 421..455
 FT /label= Integrin_alphasubunit_repeat_domain_#4
 FT Domain 478..516
 FT /label= Integrin_alphasubunit_repeat_domain_#5
 FT Domain 540..575
 FT /label= Integrin_alphasubunit_repeat_domain_#6
 FT Domain 602..640
 FT /label= Integrin_alphasubunit_repeat_domain_#7
 FT Domain 1142..1164
 FT /label= Transmembrane_domain
 FT Domain 1165..1188
 FT /label= Cytoplasmic_domain
 XX
 XX WO200073339-A1.

Search completed: April 6, 2005, 12:22:28
Job time : 5.42639 secs

11 Page Blank (uspto)

Result No.	Score	Query Match	Length	DB	ID	Description
1	192	100.0	1167	16	US-10-741-601-531	Sequence 531, Appl
2	192	100.0	1177	16	US-10-741-601-532	Sequence 532, Appl
3	69	35.9	1188	15	US-10-291-265-338	Sequence 338, Appl
4	69	35.9	1188	15	US-10-291-265-810	Sequence 810, Appl
5	59	30.7	1034	10	US-09-984-130-43	Sequence 43, Appl
6	59	30.7	1034	10	US-09-836-353A-43	Sequence 43, Appl
7	59	30.7	1120	15	US-10-262-839-6	Sequence 6, Appl
8	59	30.7	1189	10	US-09-984-130-35	Sequence 35, Appl
9	59	30.7	1189	10	US-09-836-353A-35	Sequence 35, Appl
10	59	30.7	1189	15	US-10-262-839-4	Sequence 4, Appl
11	57.5	29.9	3338	14	US-10-156-761-8464	Sequence 8464, Ap
12	56.5	29.4	1126	16	US-10-437-963-182396	Sequence 182396,
13	56	29.2	266	13	US-10-087-192-570	Sequence 570, App

```
; FILE REFERENCE: C1001500
; CURRENT APPLICATION NUMBER: US/10/741.601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-532

Query Match      100.0%; Score 192; DB 16; Length 1177;
Best Local Similarity 100.0%; Pred. No. 4.1e-17;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEVHPYGTLP--VGPGEPEKTTLRVQNLGCVVVSGLI 35
Db 952 YEVHPYGTLPVGPGEPEKTTLRVQNLGCVVVSGLI 986

RESULT 3
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291.265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338

Query Match      35.9%; Score 69; DB 15; Length 1188;
Best Local Similarity 40.5%; Pred. No. 2.1;
Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;

Qy 1 YEVHPYGTLP--VGPGEPEKTTLRVQNLGCVVVSGLI 35
Db 968 YEVKPNSSLERYDGIPPFSCIFRIQNLGLFPIHGMM 1004

RESULT 4
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291.265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03

; FILE REFERENCE: C1001500
; CURRENT APPLICATION NUMBER: US/10/741.601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-532

Query Match      100.0%; Score 192; DB 16; Length 1177;
Best Local Similarity 100.0%; Pred. No. 4.1e-17;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEVHPYGTLP--VGPGEPEKTTLRVQNLGCVVVSGLI 35
Db 952 YEVHPYGTLPVGPGEPEKTTLRVQNLGCVVVSGLI 986

RESULT 5
US-09-984-130-43
; Sequence 43, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984.130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-43

Query Match      30.7%; Score 59; DB 10; Length 1034;
Best Local Similarity 41.7%; Pred. No. 42;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGPEPKTTLRVQNLGCVVVSGLI 35
Db 981 GIGPPFSCIFRIQNLGLFPIHGIM 1004

RESULT 6
US-09-836-353A-43
; Sequence 43, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836.353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-43

Query Match      30.7%; Score 59; DB 10; Length 1034;
Best Local Similarity 41.7%; Pred. No. 42;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GPGPEFKTTLRVQNLGCVVWSGLI 35
Db 981 GIGPPFSCIFRIQNLGLFPIHGIM 1004

RESULT 7
US-10-262-839-6
; Sequence 6, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367

; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 6
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-6

Query Match      30.7%; Score 59; DB 15; Length 1120;
Best Local Similarity 41.7%; Pred. No. 46;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GPGPEFKTTLRVQNLGCVVWSGLI 35
Db 912 GIGPPFSCIFRIQNLGLFPIHGIM 935

RESULT 8
US-09-984-130-35
; Sequence 35, Application US/09984130
; Publication No. US2003005231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-35

Query Match      30.7%; Score 59; DB 10; Length 1189;
Best Local Similarity 41.7%; Pred. No. 49;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GPGPEFKTTLRVQNLGCVVWSGLI 35
Db 981 GIGPPFSCIFRIQNLGLFPIHGIM 1004

RESULT 9
US-09-836-353A-35
; Sequence 35, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
```

```
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-836-353A-35

Query Match      30.7%; Score 59; DB 10; Length 1189;
Best Local Similarity 41.7%; Pred. No. 49;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 12 GPGPEFKTTLRVQNLGCVVVSGLI 35
Db 981 GIGPFFSCIFRIQNLGLFPIHGIM 1004

RESULT 10
US-10-262-839-4
; Sequence 4, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterson, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytke, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerkhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 4
; LENGTH: 1189
```

```
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-262-839-4

Query Match      30.7%; Score 59; DB 15; Length 1189;
Best Local Similarity 41.7%; Pred. No. 49;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 12 GPGPEFKTTLRVQNLGCVVVSGLI 35
Db 981 GIGPFFSCIFRIQNLGLFPIHGIM 1004

RESULT 11
US-10-156-761-8464
; Sequence 8464, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8464
; LENGTH: 3338
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8464

Query Match      29.9%; Score 57.5; DB 14; Length 3338;
Best Local Similarity 56.5%; Pred. No. 2.3e+02;
Matches 13; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 8 TLPVGPPEFKTTLRVQNLGCVV 30
Db 2091 TLP-GPDPFSPALRLRGLGLQV 2112

RESULT 12
US-10-437-963-182396
; Sequence 182396, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 182396
; LENGTH: 1126
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 1.49374 Seconds
(without alignments)
1749.117 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEHVPVGLTVGPGPEKTKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database :
- 1: Issued Patents AA:*
 - 2: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
 - 1: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
 - 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
 - 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
 - 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	30.7	1217	4	US-09-949-016-7892
2	52.5	27.3	469	4	US-09-270-767-33880
3	52.5	27.3	469	4	US-09-270-767-49097
4	52	27.1	750	3	US-08-202-841A-2
5	51	26.6	846	1	US-08-356-354-2
6	51	26.6	846	2	US-08-778-656-2
7	51	26.6	846	4	US-09-376-045-2
8	51	26.6	890	1	US-08-145-006C-5
9	51	26.6	890	4	US-09-546-013-18
10	51	26.6	890	5	PCT-US94-00545-5
11	51	26.6	908	1	US-08-356-354-6
12	51	26.6	908	2	US-08-778-656-6
13	51	26.6	908	4	US-09-376-045-6
14	51	26.6	1053	4	US-09-394-272-6
15	51	26.6	1054	1	US-08-356-354-4
16	51	26.6	1054	2	US-08-778-656-4
17	51	26.6	1054	4	US-09-376-045-4
18	50	26.0	738	3	US-09-450-209-14
19	50	26.0	738	4	US-08-989-385-1
20	50	26.0	738	4	US-09-593-826-1
21	50	26.0	738	4	US-09-949-016-6020
22	50	26.0	753	4	US-09-949-016-11542
23	49.5	25.8	305	4	US-09-902-540-15167
24	49	25.5	1109	4	US-09-529-239D-31
25	48	25.0	51	4	US-09-621-976-7129
26	48	25.0	138	4	US-09-902-540-15376
27	48	25.0	211	4	US-09-248-796A-16892

28	48	25.0	217	4	US-09-252-991A-24818	Sequence 24818, A
29	48	25.0	266	4	US-09-248-796A-18478	Sequence 18478, A
30	48	25.0	668	3	US-09-697-367-2	Sequence 2, Appli
31	48	25.0	668	4	US-09-918-909A-2	Sequence 2, Appli
32	48	25.0	1261	4	US-09-248-796A-16620	Sequence 16620, A
33	48	25.0	1289	1	US-07-876-280-4	Sequence 4, Appli
34	48	25.0	1289	1	US-07-675-772-4	Sequence 4, Appli
35	48	25.0	1289	1	US-08-063-170-4	Sequence 4, Appli
36	48	25.0	1289	1	US-08-158-232-4	Sequence 4, Appli
37	48	25.0	1289	1	US-08-304-626-4	Sequence 4, Appli
38	48	25.0	1289	1	US-08-316-301A-4	Sequence 4, Appli
39	48	25.0	1289	2	US-08-611-928-4	Sequence 4, Appli
40	48	25.0	1289	3	US-09-173-891-4	Sequence 4, Appli
41	48	25.0	1289	3	US-09-076-137-4	Sequence 4, Appli
42	48	25.0	1289	4	US-09-738-363-4	Sequence 4, Appli
43	48	25.0	1289	5	PCT-US92-03624-4	Sequence 4, Appli
44	48	25.0	1289	6	5281530-3	Patent No. 5281530
45	48	25.0	1289	6	5426049-4	Patent No. 5426049

ALIGNMENTS

RESULT 1

US-09-949-016-7892
; Sequence 7892, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7892
; LENGTH: 1217
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7892

Query Match 30.7%; Score 59; DB 4; Length 1217;
Best Local Similarity 41.7%; Pred. No. 7.2;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGPEKTKTLRVQNLGCVVSGLI 35

Db 1010 GIGPFSCIFRIQNLGLFPIHGMW 1033

RESULT 2

US-09-270-767-33880
; Sequence 33880, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33880
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Drosophila melanogaster

```
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-33880

Query Match      27.3%; Score 52.5; DB 4; Length 469;
Best Local Similarity 34.4%; Pred. No. 21;
Matches 11; Conservative 6; Mismatches 6; Indels 9; Gaps 2;

Qy 1 YEVHPYGTLPVG-----PQPEFKTTLRVQNLGC 28
Db 174 YLHPFSLPIGSRKPGPK-----RCREFXC 200

RESULT 3
US-09-270-767-49097
; Sequence 49097, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49097
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-49097

Query Match      27.3%; Score 52.5; DB 4; Length 469;
Best Local Similarity 34.4%; Pred. No. 21;
Matches 11; Conservative 6; Mismatches 6; Indels 9; Gaps 2;

Qy 1 YEVHPYGTLPVG-----PQPEFKTTLRVQNLGC 28
Db 174 YLHPFSLPIGSRKPGPK-----RCREFXC 200

RESULT 4
US-08-202-841A-2
; Sequence 2, Application US/08202841A
; Patent No. 6218596
; GENERAL INFORMATION:
; APPLICANT: Hughes, Stephen H.
; APPLICANT: Suttrave, Pramod
; APPLICANT: Pursel, Vernon
; TITLE OF INVENTION: Enhancement of Musculature in Animals
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,841A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,415
; FILING DATE: 03-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,449
```

```
; FILING DATE: 02-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/373,864
; FILING DATE: 30-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Alicea, Hector A.
; REGISTRATION NUMBER: 40,891
; REFERENCE/DOCKET NUMBER: 015280-170300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 373
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Xaa = Trp in c-ski;
; OTHER INFORMATION: Xaa = Arg in v-ski"
US-08-202-841A-2

Query Match      27.1%; Score 52; DB 3; Length 750;
Best Local Similarity 35.3%; Pred. No. 44;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGPEF-----KTTLRVQNLGCYVVG 33
Db 67 PVMPGPFMPSDRSTERCETILEGTISCFVVG 100

RESULT 5
US-08-356-354-2
; Sequence 2, Application US/08356354
; Patent No. 5767365
; GENERAL INFORMATION:
; APPLICANT: SONNEWALD, Uwe
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
; PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,354
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/EP93/01605
; FILING DATE: 22-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42 20 758.4
; FILING DATE: 24-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
```

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 846 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-356-354-2

Query Match 26.6%; Score 51; DB 1; Length 846;

Best Local Similarity 35.5%; Pred. No. 71;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEYHPYGTLPVGPGEFKTTLRVNLCYV 31

Db 704 FKVKPGTVP--PSKELKVMRIQALRCHAV 732

RESULT 6

US-08-778-656-2

; Sequence 2, Application US/08778656

; Patent No. 5976869

; GENERAL INFORMATION:

; APPLICANT: SONNEWALD, Uwe

; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE

; PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen

; STREET: 1180 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: US

; ZIP: 10036-8403

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/778,656

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/356,354

; FILING DATE: 20-DEC-1994

; APPLICATION NUMBER: US PCT/EP93/01605

; FILING DATE: 22-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: DE P42 20 758.4

; FILING DATE: 24-JUN-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Meilman, Edward A.

; REGISTRATION NUMBER: 24,735

; REFERENCE/DOCKET NUMBER: P/951-105

; TELEPHONE: (212) 382-0700

; TELEFAX: (212) 382-0888

; TELEX: 236925

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 846 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-778-656-2

Query Match 26.6%; Score 51; DB 2; Length 846;

Best Local Similarity 35.5%; Pred. No. 71;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEYHPYGTLPVGPGEFKTTLRVNLCYV 31

Db 704 FKVKPGTVP--PSKELKVMRIQALRCHAV 732

RESULT 7

US-09-376-045-2

; Sequence 2, Application US/09376045

; Patent No. 6723898

; GENERAL INFORMATION:

; APPLICANT: Aventis CropScience GmbH

; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF

; PLANTS WITH CHANGED SUCROSE CONCENTRATION

; FILE REFERENCE: 514413-3772

; CURRENT APPLICATION NUMBER: US/09/376,045

; CURRENT FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 2

; LENGTH: 846

; TYPE: PRT

; ORGANISM: Solanum tuberosum

US-09-376-045-2

Query Match 26.6%; Score 51; DB 4; Length 846;

Best Local Similarity 35.5%; Pred. No. 71;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEYHPYGTLPVGPGEFKTTLRVNLCYV 31

Db 704 FKVKPGTVP--PSKELKVMRIQALRCHAV 732

RESULT 8

US-08-145-006C-5

; Sequence 5, Application US/08145006C

; Patent No. 5656452

; GENERAL INFORMATION:

; APPLICANT: Rao, Anjana

; APPLICANT: Hogan, Patrick Gerald

; APPLICANT: McCaffrey, Patricia

; APPLICANT: Jain, Jugnu

; TITLE OF INVENTION: NF-ATp, A T LYMPHOCYTE

; TITLE OF INVENTION: DNA-BINDING PROTEIN

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 502 or 55SX

; OPERATING SYSTEM: MS-DOS (Version 5.0)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/145,006C

; FILING DATE: October 29, 1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/017,052

; FILING DATE: February 11, 1993

; APPLICATION NUMBER: 08/006,067

; FILING DATE: January 15, 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Fraser, Janis K.

; REGISTRATION NUMBER: 34,819

; REFERENCE/DOCKET NUMBER: 04590/007001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 890
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-145-006C-5

Query Match 26.6%; Score 51; DB 1; Length 890;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 HPGTLPVGP 15
Db 805 HPLGTCPLVGP 816

RESULT 9
US-09-546-013-18
; Sequence 18, Application US/09546013
; Patent No. 6610504
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Shen
; TITLE OF INVENTION: METHODS FOR ASSAYING S-ADENOSYLMETHIONINE-DEPENDENT Methyltransferase
; FILE REFERENCE: 1937-1652
; CURRENT APPLICATION NUMBER: US/09/546,013
; CURRENT FILING DATE: 2000-04-10
; EARLIER APPLICATION NUMBER: 09/347,878
; EARLIER FILING DATE: 1999-07-06
; EARLIER APPLICATION NUMBER: 09/457,205
; EARLIER FILING DATE: 1999-12-06
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 18
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: T lymphocyte DNA binding protein: NF-At. sub.p
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5,656,452
; PATENT FILING DATE: 1993-10-29
; PUBLICATION DATE: 1997-08-12
US-09-546-013-18

Query Match 26.6%; Score 51; DB 4; Length 890;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 HPGTLPVGP 15
Db 805 HPLGTCPLVGP 816

RESULT 10
PCT-US94-00545-5
; Sequence 5, Application PC/TUS9400545
; GENERAL INFORMATION:
; APPLICANT: Rao, Anjana
; APPLICANT: Hogan, Patrick Gerald
; APPLICANT: McCaffrey, Patricia
; APPLICANT: Jain, Jugnu
; TITLE OF INVENTION: NF-ATp, A T LYMPHOCYTE
; TITLE OF INVENTION: DNA-BINDING PROTEIN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 555X
```

```
;
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00545
; FILING DATE: 18-JAN-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/145,006
; FILING DATE: October 29, 1993
; APPLICATION NUMBER: 08/017,052
; FILING DATE: February 11, 1993
; APPLICATION NUMBER: 08/006,067
; FILING DATE: January 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 04590/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 890
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US94-00545-5

Query Match 26.6%; Score 51; DB 5; Length 890;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 HPGTLPVGP 15
Db 805 HPLGTCPLVGP 816

RESULT 11
US-08-356-354-6
; Sequence 6, Application US/08356354
; Patent No. 5767365
; GENERAL INFORMATION:
; APPLICANT: SONNEWALD, Uwe
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
; TITLE OF INVENTION: PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,354
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/EP93/01605
; FILING DATE: 22-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42 20 758.4
; FILING DATE: 24-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-105
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-356-354-6

Query Match 26.6%; Score 51; DB 1; Length 908;
Best Local Similarity 35.5%; Pred. No. 77;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVV 31
Db 726 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 12

US-08-778-656-6
Sequence 6, Application US/08778656
Patent No. 5976869

GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY

COUNTRY: US

ZIP: 10036-8403

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA: US/08/778,656

APPLICATION NUMBER: US/08/778,656

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/356,354

FILING DATE: 20-DEC-1994

APPLICATION NUMBER: US PCT/EP93/01605

FILING DATE: 22-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P42 20 758.4

FILING DATE: 24-JUN-1992

ATTORNEY/AGENT INFORMATION:

NAME: Weillman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-105

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 908 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-778-656-6

Query Match 26.6%; Score 51; DB 2; Length 908;
Best Local Similarity 35.5%; Pred. No. 77;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVV 31
Db 726 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 13

US-09-376-045-6
Sequence 6, Application US/09376045
Patent No. 6723898

GENERAL INFORMATION:

APPLICANT: Aventis CropScience GmbH

TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
PLANTS WITH CHANGED SUCROSE CONCENTRATION

FILE REFERENCE: 514413-3772

CURRENT APPLICATION NUMBER: US/09/376,045

CURRENT FILING DATE: 1999-08-17

NUMBER OF SEQ ID NOS: 6

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 6

LENGTH: 908

TYPE: PRT

ORGANISM: Solanum tuberosum

US-09-376-045-6

Query Match 26.6%; Score 51; DB 4; Length 908;
Best Local Similarity 35.5%; Pred. No. 77;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVV 31
Db 726 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 14

US-09-394-272-6
Sequence 6, Application US/09394272
Patent No. 6475588

GENERAL INFORMATION:

APPLICANT: Haigler, Candace H.

APPLICANT: Holaday, A. Scott

TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED

EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE

FILE REFERENCE: 201304/1000

CURRENT APPLICATION NUMBER: US/09/394,272

CURRENT FILING DATE: 1999-09-10

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6

LENGTH: 1053

TYPE: PRT

ORGANISM: Solanum tuberosum

US-09-394-272-6

Query Match 26.6%; Score 51; DB 4; Length 1053;
Best Local Similarity 35.5%; Pred. No. 91;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVV 31
Db 911 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 939

RESULT 15

US-08-356-354-4
Sequence 4, Application US/08356354
Patent No. 5767365

GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe

TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE

PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

Search completed: April 6, 2005, 12:24:21
Job time : 2.49374 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 39.4471 Seconds
(without alignments)
2761.097 Million cell updates/sec

Title: US-09-647-544-4
Perfect score: 5932
Sequence: 1 MEHPFVTHLPVFLVLTGLC.....GFFAHKKIPREKREKLEQ 1132

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:.*
1: Pirl:.*
2: Pirl2:.*
3: Pirl3:.*
4: Pirl4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1884	31.8	1180	2 A35854	integrin alpha-1 c
2	1891.5	31.7	1151	2 A45226	integrin alpha-1 c
3	1734.5	29.2	1170	2 I45914	integrin alpha-2 s
4	1733.5	29.2	1181	2 A33998	integrin alpha-2 c
5	1712.5	28.9	1178	2 S44142	VLA-2 protein homo
6	1115.5	18.6	1170	2 S03308	cell surface glyco
7	1105.5	18.6	1163	2 I56126	lymphocyte fuction
8	1091	18.4	1153	1 RWHU1B	cell surface glyco
9	1062.5	17.9	1163	1 RWHU1C	leukocyte surface
10	1025	17.3	1153	2 S00551	integrin alpha-E c
11	928	15.6	1179	2 A53213	integrin alpha-9 c
12	714.5	12.0	1035	2 I58409	alphaP integrin -
13	700	11.8	1039	2 JC7294	lymphocyte-Peyer's
14	666.5	11.2	1039	2 A41131	integrin alpha-4 c
15	665.5	11.2	1038	2 S06046	integrin alpha-1 -
16	655	11.0	272	2 A55348	integrin alpha-7 c
17	642	10.8	1137	2 JC5950	integrin alpha-7 c
18	637.5	10.7	1041	2 T31437	alpha-7 integrin -
19	609	10.3	1135	2 I61186	integrin alpha-3 c
20	593	10.0	1051	2 A40021	integrin alpha-v c
21	587	9.9	1044	2 T10050	integrin alpha cha
22	580	9.8	1106	2 S38783	integrin alpha-5 c
23	577.5	9.7	1048	2 A27421	integrin alpha-6 c
24	571.5	9.7	1091	2 A41543	integrin alpha-6 c
25	576	9.7	1072	2 B38457	integrin alpha-6 c
26	576	9.7	1073	2 A36429	integrin alpha-6 c
27	574	9.7	1053	2 I55534	VLA-3 alpha subuni
28	567.5	9.6	1045	2 S60571	integrin alpha v c
29	567	9.6	1044	2 S16516	integrin alpha-8 c

30	566.5	9.5	1049	2 A27079	fibronectin recept
31	564	9.5	1034	2 A36108	integrin alpha-V c
32	563.5	9.5	1053	2 S44250	integrin alpha-5 c
33	560	9.4	1051	2 A35761	cell surface glyco
34	534.5	9.0	1146	2 S40311	integrin - fruit f
35	528.5	8.9	1039	2 A34269	integrin alpha-2b
36	482.5	8.1	1037	2 A60163	glycoprotein I1b -
37	475	8.0	1394	2 A29637	position-specific
38	465.5	7.8	1226	2 S44824	F34F2.1 protein -
39	443	7.5	1139	2 S28277	hypothetical prote
40	407.5	6.9	1086	2 T18523	integrin alpha cha
41	405	6.8	1115	2 T09433	integrin alpha cha
42	405	6.8	1115	2 T09403	integrin alpha cha
43	394.5	6.7	191	2 I47230	VLA-2 protein - pi
44	362.5	6.1	764	2 I36916	glycoprotein I1b -
45	292	4.9	604	2 I36917	glycoprotein I1b -

ALIGNMENTS

RESULT 1

A35854
integrin alpha-1 chain precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 23-Oct-1990 #sequence_revision 13-Sep-1991 #text_change 09-Jul-2004
C/Accession: A35854; S11243
R/ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto, J. Cell Biol. 111, 709-720, 1990
A/Title: Molecular cloning of the rat integrin alpha-1-subunit: a receptor for laminin a.
A/Reference number: A35854; MUID:90339125; PMID:2380249
A/Accession: A35854
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1180 <IGN>
A/Cross-references: UNIPROT:P18614; GB:X52140; NID:G56493; PIDN:CAA36384.1; PID:G56494
C/Keywords: cell adhesion; cytoskeleton; transmembrane protein
F:170-345/Domain: von Willebrand factor type A repeat homology <WAA>

Query Match		31.8%;	Score 1884;	DB 2;	Length 1180;
Best Local Similarity		36.2%;	Pred. No. 1.2e-131;		
Matches		432;	Conservative 219;	Mismatches 432;	Indels 112; Gaps 25;
Qy	13	LVFLTGLCSFNLDEHHPRLFP	PGPPPEAEFGYSVLQVGGGQRWMLVCAWDPGSGDRGCD	72	
Db	19	LTVTILGFCVSNVDVKNMSF	SGPVEDMFGYVQQYENEECKWVLIGSLVGPQKARTGD	78	
Qy	73	VYRCVCGAHNAPCAKGLGDY	QLGNSHPAV-----NMHLGMSILLETDDGGFMACAPL	127	
Db	79	VYKCPVGRERAMPCKVDLP	---VNTSIENVTEIKENMTFG-STLVNPNNGGLACGPL	133	
Qy	128	WSRACGSSVFSSSGICARVD	ASFQPGSLAPTAQRCPTVMVIVLDGNSIYPWSEVOTF	187	
Db	134	YAYRCGLHLYTTGLCSVDSP	TFQVNSFAP-VQECSTQLDIVLDGNSIYPWSEVIAF	192	
Qy	188	LRRVLGKFLIDPEIQVLQV	GESPVHWSLGFRTKEEVRAAKNLSRREGRETAKQ	247	
Db	193	LNDLKRMIDGPKQTVQVIG	ENVTHFNLNKYSSTEEVLVAANKIGRQGLQTMAL	252	
Qy	248	AIWVACTEGFSQSHGSRPE	ARLLAVVTDGESHGDEELPAALKACAGRVTRYGVAVLGH	307	
Db	253	GIDTARKEAFTEARGARR	GKVMVITVDGESHONYRLKQVIQDCEDENIQRFSAIILGH	312	
Qy	308	YLRRQRDPSPFLRIRTI	ASDPDRFFNFVNTDEAALTDIVDALGDRIPLGEGSHAENESS	367	
Db	313	YNRGNLSTKFEVEIKSIA	SEPTKEHFNVSDELALVTIVKALGERIFALEATADQSAAS	372	
Qy	368	FGLBMSQIGFSTHRLKDG	ILFGMVGVADWGSVLWEGGHRLLFPPRMALEDEFPALQNH	427	
Db	373	FEMEMSGTGFSAHYSQ	WDWMLGAVGAYDWNCTVVMQKQNMVIPHNTTFFOTE--PAKQNE	430	
Qy	428	--AAYLGYSVSMLLRG	RRFLFLSGAPFRHRGKVIAPQLKDKGAVRVAQSLQEGTGSY	485	

```
Db 431 PLASYLGTVNSATIPGD-VLYIAGQPRYNHTQGVYIKM-EDGNINILQTLGEGIGSY 488
Qy 486 FGSLCPLDTRDGTQTVLLVAAPMELGPQNKETGRVVYVLVGGQSLLTLQGTLP- 541
Db 489 FGSVLTITIDKDSYTDLLVGAAPMGTKEBGKVYIYAV-NQTRFQMSLEPIRQT 547
Qy 542 -----EPQDARFGAMGALPDLNQDGFADVAGAPLEDGHHQGALYLYHG 586
Db 548 CCSSLKONSCTKENKNEPCGARFOTAAVKDLNVDFNDVIGAPLEDDHAGAVIYHG 607
Qy 587 TQSGVRPHQRIAAAMPHALSIFGRSVGRDLDDGGDLVDVAVAGQAAILSSRPV 646
Db 608 SGTIREAYAQIPSGDGTLPFGQSIHGMELNDGLDVTITIGLGAALFWARDVA 667
Qy 647 HLTPSLEVTQAISVQDRCRRQGEAVCLTAALCFQVTSRTPCRWQHQFVMTASLDE 706
Db 668 VVKVTMPFNKVNQKNCRVEGKTCINATCFHVKLSKEDSIYEAADLQIRVTLDS 727
Qy 707 WTAGARAAFGSGQRLSPRLRLSVGNVTCBQLHFHVLDTSDYLRLPVALTFTFALDNTTK 766
Db 728 LRQISRFSGTQERKIQR--NITVRESECIHSFYMLDKHDFQDSVRVTLDF---NLTD 782
Qy 767 P--QPVLNBSPTSIQKLPPSKDQDNECVTLVLQVNMDIRGSKAPFVVRGGRKV 824
Db 783 PENGVLDDALPNSVHEHIPAKDCGKERCISDLTLNVST---TEKSLLIIVKSQHDKF 838
Qy 825 LVSTTLNKRKENAYNTSLIIFSNLHLASITP-QRESPIKVECAAPSAHARLCSVGHV 883
Db 839 NVSLTVKNGDSAYNTVTVQHSNLIIFSGIEEIQKDS-----CESNQNITCRVGYPF 891
Qy 884 FQTGAKVTFLLFEPFSCSLLSQVFGKLTASSDSLERNGLTQENTACTSAIYQEPHLLF 943
Db 892 LRAGETVTKIIPQFNTHLSSENAIHLSATSDSEEPLESINDNEVNISSIPVKEVGLQF 951
Qy 944 SSBSTLHRYVHPGTLP-----VG-----PGPFKTL----- 972
Db 952 YSSASEHHISVAANETPEFINSTEDICNEINVFYTKRGHFPMPBELQLSLSPNLTD 1011
Qy 973 -----RTNNAACIVQNTPEPG-----PPVHPELOHTNLNGSNTOCQVRC 1015
Db 1012 GYPVLVPIGWSSDNVNCRPSLSDPFGINSKGKQMTSKSEVLKRGTIQDCSSTCGVATI 1071
Qy 1016 HLGOLAKG-TEVSVGLLRLVHNEFFRRRAKPSLTVVSTFELGTBEGSVLQLTASRWSES 1074
Db 1072 TCSLLPSDLQSVNVL--LLWKPTFIRAHFSSNLTLRGELKSESS-LTLLSSNRKREL 1128
Qy 1075 LLEVQVT-RPILISJNLIGSVLGGLLLLALLVCLWKLQFPFAHKKIPBEBKRE 1128
Db 1129 AIQISKDGLPGRVPLWILLISAFAGLLLLMLLILALWKIGPF---KRPLKKMEK 1180

RESULT 2
A45226
Integrin alpha-1 chain - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C/Accession: A45226
R/Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A/Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A/Reference number: A45226; MUID:93155124; PMID:8428973
A/Accession: A45226
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-1151 <BRI>
A/Cross-references: UNIPROT:P56199
A/Experimental source: hepatoblastoma cell line HepG2
A/Note: sequence extracted from NCBI backbone (NCBIP:124326)
P/142-317/Domain: von Willebrand factor type A repeat homology <VWA>

Query Match 31.7%; Score 1881.5; DB 2; Length 1151;
Best Local Similarity 35.9%; Pred. NO. 1.8e-131;
Matches 426; Conservative 220; Mismatches 423; Indels 119; Gaps 23;
```

```
Qy 23 FNLDERHPLRFPGPPEAEFCYSVLQHVGGQRMMLVGA PMDGPSPGDRRGDVYRCPCVGAH 82
Db 1 FNVDVXNSMTFSFVDMFGYTVQYENEBGKWLIGSPLVGQPKARTGDVYKCPVORGE 60
Qy 83 NAPCAKGHLDGYQLGNSHPAV-----NMHLGMSLLLETGDGGMACAPLWSRACSSVF 137
Db 61 SLPCVKLDLP---VNTSIPNVTEVKENMTFG-STLVTPNPGGFLACGPLYAVRCGHLHY 115
Qy 138 SSGICARVDASFOQOSLAPTAQRCPTYMDVTVILGDSNIIYPWSEVQTLRLRLVGLKFI 197
Db 116 TTGICSDVSTFQVANSIAP-VQECSTQLDIVILGDSNIIYPWDSVTAFANDLLKRMDI 174
Qy 198 DPQIOVGLVOYGESPVHWSLGDPRTKBEVRAAKNLSSREGRETAKAIWVACTEGF 257
Db 175 GPQTQVGIYOYQENVTHEFNLNKYSTEEVLAANKIVQGGROQTWALTGTDTARKEAF 234
Qy 258 SQSHGQRPEARLLVVVTDGESHGDBELPAALKACBAGRVTTRYGIAVLGHYLRQRDPSS 317
Db 235 TEARGARRGVKVMVITDGSNDHRLKKVIQDCEDENIQRFSAIILGSYNRGNLSTEK 294
Qy 318 FLREIRTIASDPDERPPFNTVDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGF 377
Db 295 FVEEIKSIASEPTEKHFFNVSDDELAVTIKTLGERIFALEATAQDQSAASFEMMSQTGF 354
Qy 378 STHRLKDGILFGVMGVADWGGSVLWLEGGHRLPPRMALDEFPFALONHAAVLYGVSS 437
Db 355 SAHYSQDWMVLGAVGYDMNGTVVMQKASQIIIPRNTTFNVESTKNEPLASVLYGVNS 414
Qy 438 MLARGRRRLPLSGAPRHRHGKVIAPQLKDGAVRVAQSLQGBQIGSYFGSSELCPLDTR 497
Db 415 ATASSGDVLIQAQPRVNTGVIIYRM-EDGNIKILQTLGSGQIGSYFGSILTTTIDK 473
Qy 498 DGTVDULLVAAPNPLGPONKETGRVTVYLVGQGSLLTLQGTLP----- 541
Db 474 DSNTDILLVGA PMYMGTEKEEQGVYVYAL-NQTRFQMSLEPIKQTCSSSRQHNSCTT 532
Qy 542 ---EPQDARFGAMGALPDLNQDGFADVAGAPLEDGHHQGALYLYHGTVGVRPHPAQR 598
Db 533 ENKNEPCGARFGTAIAVKDLNDGFNDIVIGAPLEDHGGAVIYHGSKTRKEVQAR 592
Qy 599 IAAASMPHALSYFGRSVGDRDLDDGGDLVDVAVAGQAAILSSRPVHLTPSLVTPQA 658
Db 593 IPSGQDGKTLKFFQSIHGMELNDGLDVTITIGLGAALFWSRDAVAVVKTWMPENK 652
Qy 659 ISVQDRCRRQGEAVCLTAALCFQVTSRTPCRWQHQFVMTASLDEWTAGARAAFDGS 718
Db 653 VNIQKNCHMEGKETVCINATVCFVKLSKEDTIYEADLQYRVTLDSLQISRSPFSGT 712
Qy 719 GQRLSPRLRLSVGNVTCBQLHFHVLDTSDYLRLPVALTFTFALDNTTKP--GPVLNBSGP 776
Db 713 QERKVR--NITVRKSECTKHSFYMLDKHDFQDSVRITLDF---NLTDPEGVPVLDLSP 767
Qy 777 TSIQKLVPPSKDCGPNNECVTLVLQVNMDIRGSKAPFVVRGGRKVLVSTTLNKRKEN 836
Db 768 NSVHEYIPPAKDCGKKEKISDLSLV-----ATTEKOLLIVRQNDKFNVSILTVKNTKDS 823
Qy 837 AYNYSLSIIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVFPQCAKVTFLLEF 896
Db 824 AYNTRTIVHYSPLNVPSGI-----EAIQKQSC--ESNNHNTCKVGYPLRREGVMVTFKILF 877
Qy 897 EFCSSLSLQVFGKLTASSDSLERNGLTQENTACTSAIYQEPHLLPSSSTLHRYBVHP 956
Db 878 QFNYSYLMENVTIYLSATSDSEPPETLSNVNVSIPVKEVGLQYSSASEVHISIAA 937
Qy 957 YGTLP-----VG-----PGPEFKTLRTNNAACIVQNTPEPGPVH 993
Db 938 NETVPEVINSTEDIGNEINIFYLIRKSGSPMPBELKLSIS-----FPNMTSNGYPVLY 990
Qy 994 PSELQHTNRLN-----GSNTQCO-----VVRCHLQOLA 1021
Db 991 FTGLSSSENANCRPHIFEDPFSINSKGKMTTSTDLHKLKRGTLDCNTKFAITCNLTLS-S 1049
```

```
Qy 1022 KGEVSVCLLRLVHNEPRRAKFKSLTVVSTFELGTEGSLVQLQTEASRWSESLLEVQOT 1081
Db 1050 DISQVNSL--ILWKPTFKSYFSLNLTIRGELSENAS-LVLSSSNQKELAIOISKD 1106
Qy 1082 -RPLISLWILGVLGGLLLALLVFLCWLKGLGFFAHKKIPEEEKREE 1128
Db 1107 GLPRVPLWILLSAFAGLLLLMLLILALWLGFF--KRPLKKMEK 1151

RESULT 3
I45914
integrin alpha 2 subunit - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
C:Accession: I45914
J:Kamata, T.; Puzon, W.; Takada, Y.
J: Biol. Chem. 269, 9659-9663, 1994
A:Title: Identification of putative ligand binding sites within the I-domain of integrin
A:Reference number: A54402; MUID:94193647; PMID:7511592
A:Accession: I45914
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1170 <RAM>
A:Cross-references: UNIPROT:P53710; GB:L25886; NID:9439695; PIDN:AAB59255.1; PID:g439696
F:161-336/Domain: von Willebrand factor type A repeat homology <WMA>

Query Match 29.2%; Score 1734.5; DB 2; Length 1170;
Best Local Similarity 33.8%; Pred. No. 1.6e-120;
Matches 405; Conservative 229; Mismatches 456; Indels 109; Gaps 29;

Qy 9 LFLPLVLTGL---CSFNLDEHPRLPFGPEAFGYSVLQHVGGGQRMWLVGAPWDGP 65
Db 2 LQLVLVFSQGLNCCVAYNVLGPKAKIFSGSPSQFGYAVQQFNPKNWLLVSGSPMSGF 61
Qy 66 SGDRRGDVYRCPVCGAHNAPCAKGHIL-GDYOLGNSHHPAVNMHLCMSLLETGDDGFMAC 124
Db 62 PKRMGDVYKCPV-DLSTTTCENLQSTSMNSVNTKNTNSLGLTLTRNVGTGGFLTC 120

Qy 125 APLMSRACGSSVFSGICARVDASFPQGSIAPTAQRPTMYDVVIVLDGNSNTPMSEV 184
Db 121 GPLVAQQCGSYTTGVCSDVSPDFQLRTSPAPAVQTCPSFIDVVVWCDSNSIYPMDAV 180
Qy 185 QTFRLRLVGLFIDPEQIOVLQVYGESPVHWSLGDPRTEEVVRAAKNLSRREGRETK 244
Db 181 KNFLEKPVQGDIDGPTKTQMLGIOYANPNRVFVFNLTNTPKSKDEMFKATQSTFOYCGGLDN 240
Qy 245 TAOAIMVACTGFGSOSHGRPEARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIATV 304
Db 241 TFKAIOYARDATYSTAGRGFGATKVMVVVTDGSHDGSGLKXVIDQCCKNLIIRFLGIATV 300
Qy 305 LGHYLRQRDPSSFLREIRTIASDPDERFFNVNTEAALTDIVDALGDRIFLGEBSHAEN 364
Db 301 LGYLNRLALDTKNLKEIKAIASIPTEHFNVSDEADLLEKAGTIGSQIFSIEGT-VQG 359
Qy 365 ESSFGLEMSQGFSTHRLKDG--ILFGMVGAYDGGSVLW-LEGHRLFPFRMALEDEPP 421
Db 360 GDNFQEMSGVQFSAEYSPQNMLMGAVGAYDWSGTVQKTPRGHLIFS-----KQAFE 414
Qy 422 PALO--NHAAYLVGSVSMMLRGRRFLSGAPFRHRGKVIATFOLKXGDGAVRVAQSLQG 479
Db 415 QILQORHNSYLGYSVAS-ISTGNSVHFVAGAPRANYTGQIVLVSVNENGNTVIQSQRG 473
Qy 480 EQIGSYFSGELCPDTRDGTDTDLLVAAPMFLGQPNKETGRVVVYLVGQOSSLTLQOTL 539
Db 474 DQIGSYFSGVLCAVDVKNKDTITDLLVCGAPMYMNDLKEEGRVYVLTITKILANWHQPLE 533
Qy 540 QPEPPQDARFGAMGALPDNLQDGPADVAVGAPLEDGHQGLALYLYHGTQSGVRPHPAQRI 599
Db 534 GPNGLENARFGSAALSDINMDGNDVIVGSPLENQNSGAVIYVNGHEGMIRLRYSQKI 593
Qy 600 AAASMPHA--LSYFGRSVDGRDLDDGDLVDVAVGACQAAILLSRPVHLTPSLVTPQ 657
Db 594 LGSDFRASHLQYFGRSLDGVDLNGDSITDVSVGAFQGVVOLWSQSADVSVDASFTPK 653
```

```
Qy 658 AISVVQDRCRRRQGEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDWETAG----ARA 713
Db 654 KITLLNKNAE-----IKLCLFSAKFR-PTNQNNQVAIVYNTITDEDOFSSRVISRG 704
Qy 714 AFDGSGORLSRRLRLSVGNVTCBQLHFHVLDTSDYLRLPVVALTVTFALDNTTKPG--PVL 771
Db 705 LFKENNERCIQKTMIVSQAO-RCSEYIIHQPSDDIISPLMLCNWISLEN---PGTNPAL 760
Qy 772 NEGSPTSIOQLVPFSCDGPDNECVTDLVQLQNMNDRGSRKAPFVVRGGRKVLVSTTLE 831
Db 761 EAYSETVKVFSIPPHKCGDGDGVCISDLVLNV-QQLPATQCOQPFIVSNQNKRLTFSVQLK 819
Qy 832 NRKENAVNTSLSIIFSRLHLASLTPORESPI-----KYEC-AAPSAHARLCSVGHVPFTG 887
Db 820 NKESAYNTEIVVDSENLFFASWS-----MPVDGTEVTCQIASQKSVTCNVGYPALKSK 875
Qy 888 AKVTFLEFFSCSLLSVOFGKLTASSDSLSRLNGTLQENTAQTSAYIOYEPHLLFSSES 947
Db 876 QQVTFITFNFNQLQNLQNASISFRALSESQEN--MADNSVNLKLSLLYDAEHTITRST 933
Qy 948 TLHRYEVHPYGTL-----PVGPGPEFKTTLRTNN-----ASCIVQ--NLTPPPGP-- 990
Db 934 NINFEVSLDGNVSSVHVSFEDIGPKTFISIKVTGSPVPSMASVIIHIPOYTKDKNELM 993
Qy 991 -----PVHP-----BELQHTNRLNGSNTOCQVVRCHLQ 1019
Db 994 YLTGVHTDQAGDISCEAEINPLKITQTSVSSVSFKSENFRHIKELNCRKTASCNIMCWLRD 1053
Qy 1020 LAKGTEVSVGLLRLVHNEFFERRAKFKSLTVVSTFELGTEGSLVQLQTEASRWSESLLEV 1079
Db 1054 LQVKGTEVFLNVSTRINWGTFAASTFTQVLTAAAEIDTNPQIYVIBE-----NTVTPLT 1109
Qy 1080 QTRP---ILISLWILGSLVGLLLALLVFLCWLKGLGFFAHK-----KIPPEEKREEKL 1130
Db 1110 IMKPEKVEVPTGVIVGSVIAGILLALLVALLKWLKGLGFFKYEKMAKNPDETDETTTEL 1168

RESULT 4
A33998
integrin alpha-2 chain precursor - human
N:Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 ch
C:Species: Homo sapiens (man)
C:Date: 30-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 09-Jul-2004
A:Accession: A33998; B56793; A53117
R:Takada, Y.; Hemler, M.E.
J: Cell Biol. 109, 397-407, 1989
A:Title: The primary structure of the VLA-2/collagen receptor alpha(2)-subunit (platelet
A:Reference number: A33998; MUID:89308879; PMID:2545729
A:Accession: A33998
A:Molecule type: mRNA
A:Residues: 1-1181 <TAK>
A:Cross-references: UNIPROT:P17301; GB:X17033; NID:g33906; PIDN:CAA34894.1; PID:g33907
A:Note: The authors translated the codon GAT for residue 802 as Gln, GTC for residue 803
R:Caicmel, B.; Parentier, S.; Leung, L.L.; McGregor, J.L.
Biochem. J. 279, 419-425, 1991
A:Title: Separation of important new platelet glycoproteins (GPIa, GPIC, GPIIb, GPIIc and GPIIb
A:Reference number: A56793; MUID:92061944; PMID:1953640
A:Accession: B56793
A:Molecule type: protein
A:Residues: 30-43 <CAR>
R:Experimental source: platelet
R:Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Teung, Y.L.; Gafford, A.
J. Biol. Chem. 269, 463-469, 1994
A:Title: The human alpha-2 integrin gene promoter. Identification of positive and negati
A:Reference number: A53117; MUID:94103255; PMID:8276836
A:Accession: A53117
A:Molecule type: DNA
A:Residues: 1-16, 'V', 18-21 <ZUT>
A:Cross-references: GB:L24121; NID:g400342; PIDN:AAA16619.2; PID:g4583535
A:Note: authors translated the codon GTA for residue 17 as Leu
C:Genetics:
A:Gene: ITGA2; CD49B
```

A/Cross-references: GDB:128031; OMIM:192974

A/Map position: Sqll.1-1-Sqll.2

C/Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein

F/1-29/Domain: signal sequence #status predicted <SIG>

F/30-1133/Domain: extracellular #status predicted <EXT>

F/172-347/Domain: von Willebrand factor type A repeat homology <WAA>

F/1134-1154/Domain: transmembrane #status predicted <TM>

F/1155-1181/Domain: intracellular #status predicted <CYT>

F/1105,112,343,432,460,475,699,1057,1074,1081/Binding site: carbohydrate (Aen) (covalent)

Query Match 29,24; Score 1733,5; DB 2; Length 1181;
Best Local Similarity 33,74; Pred. No. 2e-120;
Matches 410; Conservative 211; Mismatches 431; Indels 159; Gaps 32;

Qy 11 LPLVFLTGL-----CSPNLDHHPRLFPGPPEAEFGYSVLQHVGGQRMVLVGAPWD 63
Db 11 LPLLLVLAISQILNCCLAYNGLPEAKIFSGPSSEQFGYAVQFIPKGNWLLVGPWS 70

Qy 64 GPSGDRGDVYRCPVGAHNAPCAKGLH-GDYQLGNSHHPANMHLGMSLLETDDGGPFM 122
Db 71 GFPENRMGDVYKCV-DLSTATCEKLNLTSTISPNVTEKMTNMSLGLILTRNMGTOGFL 129

Qy 123 ACAPLMSRACGSSVFSFGICARVDASFPQGS LAPTAQRCPYMDVIVLVDGNSIYPWS 182
Db 130 TCGPLMAQQCGNQYITGVCSIDISDFQLSASFSPATQPCPSLIDVVVVCDESNIYPWD 189

Qy 183 EVQTFRLRLVGLPIDPEIQVLGVQGESPVHWSLGDPRKTEEVRAAKNLSRREGRE 242
Db 190 AVKNFLSEKVFQGLDIGTKTQVGLIQVANNPRVFNLTYYTKKEEMIVATSTQSYGGDL 249

Qy 243 TKTAQAIMVACTBEGFSOSHGRPEARLLVVVTDGESHGDBEELPAALCACBAGRVTYGI 302
Db 250 TNTFGAIQYARKYAYSAAGRRRSATKVMVMTDGS HDGSMKKAVIDQCNDHNLRFGI 309

Qy 303 AVLGHYLRQRDPSPFLREITIASDPPERFPFNVTDEALTIDVLDGRIFGLESHA 362
Db 310 AVLGYLNRNALDTKNLIKELKAIASITERYFFNVSDAALLKAGLTGEGIFSEGT-V 368

Qy 363 ENESSFGLEMSQIGFSN--HRLKGILFGMVGADWGGSVLW-LEGGHRLPPRMALEDE 419
Db 369 QGGDNFQWMSQVGFSDYSSQNDILMLGAVGAFGWSGTIVQKTSHGHLIFP-----KQA 423

Qy 420 FPPALQ--NHAAYLGYSVSSMLRGRRFLFSGAPRFRHRGKVIAPQLKDGAVRVQAQL 477
Db 424 FQQLQDRNHSSYLGYSVAA-ISTGESHTFVAGAPRANYTGQIVLYSVNENGNTVIQAH 482

Qy 478 QGEQIGSYFGSELCPETDRDGTVDLLVAAPMFLGPONKETGRVYV-----LVGQSL 532
Db 483 RGDQIGSYFGSVLCSVDVDKDTITDVLVAGAPMYMSDLKKBEGRVLYFTIKKGLGQH 542

Qy 533 LTLGTLQPPQDARFGFAMGALPDNLNDGFADVAVGAPLEDHGQALYLYHGTQSQRV 592
Db 543 --LEG---PEGIENTRFGSAALASLDINMGDFNVI VGSPLENQNSGAVIYNGHQGTIR 597

Qy 593 PHPAQRATAA--SMPHALSYFGRSVDGRDLDDVDVAVGAQQAAILLSRPIVHLTP 650
Db 598 TKYSQKILGSDGARFSLHQLYFGRSLDGVGLNGDSITDVSIGAFQGVVQLMSQSIADVAI 657

Qy 651 SLEVTPOAISVQRDCRRRGOEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDIEWTAG 710
Db 658 EASPTPEKITLVNKNQAQ-----IILKLCFSAKFR-PTKQNNQVAIVVYNITLDA---- 704

Qy 711 ARAAPDGGQRLSPRL-----RLSVGNV-----TCEQLHFHVLDTSYLRLPVALTVT 758
Db 705 -----DGFSSRVTSRGLFKENNERCLQKNMVVNAQSCPEHIYIQEPSDVVNLSLDRVD 759

Qy 759 FALDNTTKPG--PVLNIGSPSTSIQKLVPPFSKDCPDNECVTDVLQVNMDIRGSRKAPFV 816
Db 760 ISLEN--PGTSPALAEVSETAKVPSIPFHKDCGEDGLCISDLVLDVR-QIPAAQEPFI 815

Qy 817 VRGRRKVLVSTTLNRKKNAYNTSLSIIFGRNLHLASLTPORESPI---KVEC-AAPSA 872
Db 816 VSNQNKLLFTSVTLNKRKESAYNTGIVVDFFENIFFASFS----LPVDGTETVQVAASQ 871

Qy 873 HARLCSVGHVVFOTGAKVTPLLBFEPSCSLLSQVFGKLTASSDSLRNGTTLQENTFAQTS 932

Db 872 KSVACDVGYPALAKREQQVTTINFDNLQNLQNAQSLFQALSESQENKA--DNLVNLK 929

Qy 933 AYIQYEPHLLFSSSESTLHRYEVHPYGTLP-----VQGPGEFKTTLRTNNAACIVQ----- 982

Db 930 IPLYDAEIIHLTRSTNINFYEISDGNVPISVHSFEDVGPKFIFSLKVTVTGSSVPVSMATV 989

Qy 983 ----NUTEPPGP-----PVHP-----EELQHTNRLN 1004

Db 990 IIHIPQYTKENPLMYLTGVQTDKAGDISCNADINPLKIQTSSTSSVFKSENFRTKELN 1049

Qy 1005 GSNTQCOVVRCHLGQLAGKTEVSGLLRVLVHNEFFRRAKPKSLTVVSTFELGTBEGSVLQ 1064

Db 1050 CRTASCSNVTCWLKDVHMKGEYFNVVTRVWNGTFFASSTPQTQVLTAAABINTYNPBY- 1108

Qy 1065 LTEASRWSESLLEVQVTRPLISLWI-----LIGSVLGSLMLALLVPLCLWK 1111

Db 1109 -----VIDNTVTIPLMIKMPDEKAEVPTGVTIGSIAGILLULLALVAILWK 1155

Qy 1112 LGFFAHK-----KIPSE 1123

Db 1156 LGFPKRYEKMTKNPDE 1172

RESULT 5

S44142

VLA-2 protein homolog - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004

C:Accession: S44142

R:Edelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, submitted to the EMBL Data Library, January 1994

A:Description: The mouse VLA-2 homologue supports collagen and laminin adhesion

A:Reference number: S44142

A:Accession: S44142

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1178 <EDE>

A:Cross-references: UNIPROT:Q62469; EMBL:Z29987; NID:g473098; PIDN:CAA82877.1; F:169-344/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 28.98; Score 1712.5; DB 2; Length 1178;

Best Local Similarity 34.78; Pred. No. 7.3e-119;

Matches 412; Conservative 217; Mismatches 457; Indels 103; Gaps 32;

Qy 9 LFLPLVLPLT--GL--CSPFNLDHHPRLFPQPPRAEFGYSVLQHVGGQRWMLVCAWPMDGP 65

Db 10 LLLQLMLVQGLINCLAYNGLPKATFSPGSEQFGYSVQQLTNPQGNWLLVGSPPSGF 69

Qy 66 SGDRRGDVYRCYPVGGAHNAPCAKHGL--GDYQLGNSSHPAVNMHGLMSLLETDDGGGPMAC 124

Db 70 PENRMGDVYKCPV--DLPTATCEKLNQLNSASISNVTEIKTNMSLGLTLTRNPGTGGPLTC 128

Qy 125 APLWSRACGSSVSSGICARVDASFPQGSIAPTAQRCPYMDVIVLDCSNSIYPMSEV 184

Db 129 GPLMAHCGNQYATGICSDVSPDFQLTFSFAVQACPSLDVWVVCDSNSIYPMSEV 188

Qy 185 QTFLRRRLVKGKLFIDPEQIQVGLVQYGSPPVHSESLGDFRTKEEVRAAKNLSRRGRETQ 244

Db 189 KNFLVKFTVGLDIGPKTQVALIQYANEPRIIFNLNDFETKEDMVQATSETROHGGDLTN 248

Qy 245 TQAQIMVACTGEPFSGHGRPEARLLVVYTDGESHGGBELPAAKACAGRVTRYGIIV 304

Db 249 TFRAIREFARDYAYVQTSGRPGATKVMVWVYTDGESHGSKLKTIVIQCNDDIELRFGIIV 308

Qy 305 LGHYLRQRDPSPFLRIITIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGSHAEN 364

Db 309 LGYLNRLALDTKNLIKELKAIATPTERYIFNVADAELLEKAGTLGEQIFSTIEGT--VQG 367

Qy 365 ESSFGLMSQIGFSTHRL--KDGILFMGVAYDWGGSVLBLEGHR--LFPFPRMALEDEF 420

Db 368 GDNFQMEQAQVGSADYAPQNDILMLGAVGAFDMSGTILV-QETSHKPVIPP-----KQAF 421
Qy 421 PPALQ--NHAAYLYGSVSSMLLRGRRLFLSGAPRRHRGKVIAPOLKKGAVVAQSLQ 478
Db 422 DQVLQDRNHSFLGSVAAISTEDGVH-FVAGAPRANVTGQIVLYSVNKGQNVTVIQSHR 480
Qy 479 GEQIGSYFGSELCPDTRDGTDTDLVAAPMFLGPQNKETGRVYVVLVGGQSLLTQGT 538
Db 481 GDQIGSYFGSVLCSVDVDKOTITDVLVGAFTYWNLDLKEGKGYLFTITKILNQHQFL 540
Qy 539 LQPEPPQDARFGFANGALPDINQDGFADVAVGAPLEDCHQCALYLYHGTQSGVRPHPAQR 598
Db 541 EGPEGTGNARFGSAIALSDINMGDFNDVIVGSPVENENSGAVIYNGHQTITKYSQK 600
Qy 599 IAAA--SMPHALSIFYGSRVGRDLDDGLVDVAVGAGAAILLSSRPVHLTSPLEVTP 656
Db 601 ILGNGAPRRHLQFGRSLDGYGLNGDSITDVSIGALQGVQILQWSQIADVAEALFTP 660
Qy 657 QAISVQDRCRRQGEAVCLTAALCFQVTSPTGCRWDHOFYMRFTASIDETAGARAAD 716
Db 661 DKITLLNKDAK-----ITLKCFRAEFPAQO--NNQVAILFNMTLADAGHSRVTSR 711
Qy 717 GSGQRLSPRRLR--LSVGNV--TCQLHPLHVLDTSDYLRPVALVTTFALDNTTKPG--PVL 771
Db 712 GVFRENSERFLQKNMVNVEQKSEHHSIQKPSDVNPLDLRVDISLEN---FGTSPAL 768
Qy 772 NEGSPISQIKLVPSKCGPPNECVTDLVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLE 831
Db 769 EAYSETVKVFSIPFYKEGCSGICISDILDV-QOLPAIQTSPIVSNQNKRLTFSVLK 827
Qy 832 NRKENAVNTSLIIFSRNLHLASLTPORESPI---KVECAPSAHARL--CSVGHVPVQTG 887
Db 828 NRGESANTVVLAEFSENLFPAFSF---MPVDGTEVTCVSGSQKSVCTDVGVPALKSE 893
Qy 888 AKVTFLEFEFPCSLLSQQVEKGLTASSDSLSERNGLTQENTAQTSAYTOVEPHLLFSSES 947
Db 884 QOVTTFINFDNLQNLQNAINFQAFSESQETNKA--DNSVSLTIPLYDAELHLRST 941
Qy 948 TLHRYEHPYGTLP-----VQFG-----PEFK----- 969
Db 942 NINFEYSSDENAPSVIKSVEDIGPKFTFLSVKTAGSAPVSMALVTIHIPOYTKENPLL 1001
Qy 970 --TTLRTNNA---SCI--VQNLTEP---PGPVHPELOHTNRLNGSTNQOVVRCHLQ 1019
Db 1002 YLTGIQTDQADISCTABINFLKLPHTAPSVSFKNENFRHTKELDCRTTSCNTTCLWKD 1061
Qy 1020 LAKTEVSGLRLVHNEFRRAKFSKLTVYSTFELGTEGSLVQLTEASRWSLSLEV 1079
Db 1062 LHMKAEIFINTVTRVWRTFAASTFQVQLTAAAEIDTHNPQ-LFVIEENAVTIPLIMK 1120
Qy 1080 QTRPILISLWILGSLVGLGLLLALLVFLMKLQFF--AHKKI---PEE 1123
Db 1121 PTERAEVPTGVIGSIIGIILLAMTAGLWKLGFGRKQYKMGQNPDE 1169

RESULT 6
S03308
cell surface glycoprotein Cd11a precursor - human
N;Alternate names: leukocyte adhesion glycoprotein LFA-1 alpha chain; leukocyte function
C;Species: Homo sapiens (man)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
R;Accession: S03308; A47458; A47565; A48759; S36044
R;Larson, R.S.; Corbi, A.L.; Berman, L.; Springer, T.
J. Cell Biol. 108, 703-712, 1989
A;Title: Primary structure of the leukocyte function-associated molecule-1 alpha subunit
A;Reference number: S03308; MUID:89139587; PMID:2537322
A;Accession: S03308
A;Molecule type: mRNA
A;Residues: 1-1170 <LAR>
A;Cross-references: UNIPROT:P20701; UNIPROT:Q9UBC8; EMBL:Y00796; NID:g31421; PIDN:CAA687
A;Note: part of this sequence was confirmed by protein sequencing
R;Cornwell, R.D.; Gollan, K.A.; Hickstein, D.B.
Proc. Natl. Acad. Sci. U.S.A. 90, 4221-4225, 1993

A;Title: Description of the leukocyte function-associated antigen 1 (LFA-1 or CD11a) pro
A;Reference number: A47458; MUID:93248261; PMID:8097887
A;Accession: A47458
A;Molecule type: DNA
A;Residues: 1-20 <COR>
A;Note: sequence extracted from NCBI backbone (NCBIN:130862, NCBI:P130863)
R;Shelley, C.S.; Farokhzad, O.C.; Arnaout, M.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 5364-5368, 1993
A;Title: Identification of cell-specific and developmentally regulated nuclear factors t
A;Reference number: A47565; MUID:93281759; PMID:8099450
A;Accession: A47565
A;Molecule type: DNA
A;Residues: 1-20 <SHE>
A;Cross-references: GB:M95609
R;Nueda, A.; Lopez-Cabrera, M.; Vara, A.; Corbi, A.L.
J. Biol. Chem. 268, 19305-19311, 1993
A;Title: Characterization of the CD11a (alphaL, LFA-1alpha) integrin gene promoter.
A;Reference number: A48759; MUID:93374910; PMID:8103515
A;Accession: A48759
A;Molecule type: DNA
A;Residues: 1-20 <NUE>
A;Cross-references: EMBL:Z22804; NID:g311405; PIDN:CAA80461.1; PID:g311406
C;Genetics:
A;Gene: ITGAL; CD11A
A;Cross-references: GDB:119757; OMIM:153370
A;Map position: 16p11.2-16p11.2
C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C;Keywords: cell adhesion; cytoskeleton; glycoprotein; heterodimer; surface antigen; tra
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1170/Product: leukocyte adhesion glycoprotein LFA-1 alpha chain #status predicted <
F;154-317/Domain: von Willebrand factor type A repeat homology <WAA>

Query Match 18.8%; Score 1115.5; DB 2; Length 1170;
Best Local Similarity 30.0%; Pred. No. 2.3e-74;
Matches 368; Conservative 191; Mismatches 454; Indels 213; Gaps 55;

Qy 11 LPLVFLTGL-----CSPFNLDHHPFLFPGPPEA--EFGYSVLQHVGGQRMVLVGPWD 63
Db 9 MAMALLSGFFFPASSNLDVRGARSF-SPPRAGRHFGRVQLQ-VGNG---VIVGAPGE 63
Qy 64 GPSDRDRGDVYRCVPGGAHNAAPCAKGLGDYQGLGNSHPAANMHLGMSLLETDCGGGFMA 123
Db 64 GNS---TGLYQCOSGTGCHLPVT-----LRGSNY--TSKYLGMTLATDPTDGSILA 110
Qy 124 CAPLWSRACSSVSSFGICARVDASFO-PQGS LAPTAQRC-PTVMVIVVLDGNSNYP- 180
Db 111 CDPGLSRTCDQNTYLSGLCYLFRNLQCPMLQGRPGQECIKGNVDLVFLPDGMSLSQPD 170
Qy 181 -WSEVQTFLLRLVCKLFDPEQIQVGLVOYGESPVHWSLQDF---RTKEBVVRAAKNLS 236
Db 171 EPQKILDFMKDVMKKL--SNTSYQFAAVQFSTSYKTFDFSDYVYKWDPPDALLKHVXHML 228
Qy 237 RREGRETKAAIMVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKACEAGR 296
Db 229 LL-----TNTFGAINVATEVFEELGARPDATKVLIIITDGEATDSGNIDAKD----- 278
Qy 297 VTRYGI AVLGHYLRQRDPSSFLEIRTIASDPDERFFNVNVDAAALTDIVDALGDRIFG 356
Db 279 IIRYIIGIKGHFQTKESQET-----LHKFASKPASEFVKILDTPKLDLFTLQKKIYV 333
Qy 357 LEGSHAENESSFGLMSQIGFSTHRLKDXGLFGVAYDNGCVSLWLEGGHRLFPFPMAL 416
Db 334 IEGTSKQDLTSFNMELSSSGISADLSRGHVAVGAVGADWAGGFLDKADLQ----- 385
Qy 417 EDPEF---PPALONHAAVLYGVSSVSMMLRGRRRLFLSGAPRRHRGKVIAPOLKKGAA-V 471
Db 386 DDTFIGNEPLTPEVRAGYLGTVTWLPSRQKTSLLASGAPRYQHMGRVLLFQEPQGGHW 445
Qy 472 RVAQSLQGEIGSYFGSELCPDTRDGTDTDLVAAPMFLGPQNKETGRVYVY---LVG 528
Db 446 SQVQTIHGQIGSYFGSELGCVDDVDQGETELLIGAPLFVGEQ--RGRVFIYQRQLG 503
Qy 529 QQSLLTLOGTLOPEPPQDARFGFANGALPDNLQDGFADVAVGAPLEDHGHALYHGTQ 588

```
Db 504 FEEVSELQ--DPGYPL-GRFGEAITALTDINGDGLVDVAVGAPLEE--QCAVYIFNGRH 558
Qy 589 SGVRPHPAQRTAAASMPHALSYFGRSDVDRDLDDGLDLDVAVGAQGAAILLSRPVHL 648
Db 559 GGLSPQSPQRLEGTVQLSGIQMFGRSIHGVDLEGGDLADVAVGAESQMVLSSRPVDM 618
Qy 649 TPSLEVTPOAISVVQRDC---RRRQGEAVCLTAAALCFQVTSRTPFGRWDHQFYMR----- 699
Db 619 VTLMSPSPAEIPVHEVECSYSTSNKMEGVNIT--ICFOIKSLYP-----QFGRLVANL 671
Qy 700 -FTASLDWETAGARAAAFDGSQRSLRRLRLSVGNVTCQLHFHV-LDTSYLRPVALTV 757
Db 672 TYTLQLDGHTRRRGLFPGRHSL--RRNIAVTTSMSTDFSPHFVPCVQDLSPINVSL 729
Qy 758 TPAL---DNITKPPVLNBSPTSIOKL-----VPSKDCGPDNECVTDLVLVQVN-MD 806
Db 730 NFSWESEGTDRQRAQKQIPILRPSLHSETWEIPFKNCGEDKKCBANLRVSPGAR 789
Qy 807 IRGSRKAPFVVRGRRKRVLYSTLTLENKENAYNTSLGIIFSRNHLASLTPQR-ESPIKV 865
Db 790 SRALRLTAFA-----SLSVLSLSNLEEDAYWQLDLHPFPGLSFRKVEMLKPHSQIPV 843
Qy 866 ECAAPSAHARL-----CSVGHPVQFGAKVTFLLPEFSCSSLLSQVFGKLTASS--DS 917
Db 844 SCELPBESRLLSRALSCNVSSPIFKAGHSAV-----LQMMFNTLVNSSWGDS 891
Qy 918 LERNGT-----LQENTAQTSAYIQIEPHLLF--SSESTL-----H 950
Db 892 VELHANTCNNEEDSLDENSATTIIPILYPINLIQOEDSTLVYVFTPKGPKHQVKH 951
Qy 951 RYEV-----HPYGTL-----PVGPPEKTLRTNNASCIVQNLTEPPGPPVHP 994
Db 952 MYQVRIQPSIHQNIPTLEAVGVVPQPPSECP-----ITHQWS--VQ--MEPP-VCHY 1000
Qy 995 BELQHTNRLNGSTQC---QVVRCHLQOLAKGTVEVSGLRLV---HNEPFRRAKPKSL 1047
Db 1001 EDLE---RLPDAEPCLPGLFRCVVPVFRQILVQVIGTLELVEIEBASSMFLSCSLSI 1057
Qy 1048 TVVST--PELGTSGSVLQLTEARWSESLLEVQTRPILISLWILGSLGGHLLALL 1105
Db 1058 SFNSKSHFHLVGSNASLAQVY-----MKVDVVEYKQMLY-LYVLGS--IGGULLLLLI 1107
Qy 1106 VFCLWLKGFFAHKKIPBEEKREEKLE 1131
Db 1108 FIVLYKVGF-----KGNLKERME 1126

RESULT 7
I56126
Lymphocyte fuction-associated molecule-1-alpha - mouse
C/Species: Mus musculus (house mouse)
C/Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C/Accession: I56126
R/Kaufmann, Y.; Teeng, E.; Springer, T.A.
J. Immunol. 147, 369-374, 1991
A/Title: Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit
A/Reference number: I56126; MUID:91268576; PMID:2051027
A/Accession: I56126
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1163 <RES>
A/Cross-references: UNIPROT:P24063; GB:M60778; NID:g198785; PIDN:AAA39426.1; PID:g198786
C/Genetics:
A/Gene: LPA-1
C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology
F;151-315/Domain: von Willebrand factor type A repeat homology <VMA1>

Query Match 18.6%; Score 1105.5; DB 2; Length 1163;
Best Local Similarity 28.8%; Pred. No. 1.3e-73;
Matches 359; Conservative 180; Mismatches 449; Indels 259; Gaps 49;
Qy 12 PLVFLTGL-----CSPFNLDHHPRLPFGPPEAFGYSVLQHVGGQRWMLVGAPWDGSPS 66
```

```
Db 8 PRLLLLQLFLAKAWSYNLDTRPTQSFQAQGRHFGVQLQIEDG---VVVGAPE--- 60
Qy 67 GDRRGDVYRCPVCGAHNAPCAKHLGDYQYLGNSHSHPAVNMHLGMSLLETDCGGPACAP 126
Db 61 GDMWTGGLYHCRTSSEFCQPVVS-----LHGSNH--TSKILGMLTATDAKSSLLACDP 110
Qy 127 LWSRACSSSVFSSGICARVDASFPQGS LAPTAQRCPY-----MDVVIVLDGNSI- 178
Db 111 GLSRTCDQNTYLSGLC-----YLPQSLGPMQNRPAQECMKGVLDVLPFDGQSOLD 165
Qy 179 -YPMSEVQTFRLRLVGLKFLIDPQIQVGLVQYGESPVHWSLGDFTKKEVVRRAAKNLSR 237
Db 166 RKDFEKILEPMKDMVKRL--SNTSYQFAAVQFSDCTCRTEFTFLDY-----VKQNKNDV 217
Qy 238 REGRE-----TKTAQAINVACTEGESQSHGGGPEARLLVVTVDGESHGDEELPAALKA 291
Db 218 LLSGVQPMFLTNTFRAINVYVAHVFKBESGARPDATKVLVITDGEASDKGNISA---- 273
Qy 292 CEAGRVTYGIADVGLHYLRQRDPSPSFLREIRTIASDPDERFFNFVTDAAALTDIVDALG 351
Db 274 --AHDITRYIIGIKHFSVKQ-----KTLHIFASEPVVEFVKILDTPEKLDLFDLQ 326
Qy 352 DRIFLEGSHAENESSFGLEMSQIGFSTHRLKDGILFGVMGAYDMGGSVLWLEGHRLFP 411
Db 327 RRIYATEGTRQDLTSFMELSSGISADLSKGHAVVGAAGKDWAGGFLDLR----- 379
Qy 412 PRMALED-----EPFPALQNHAAVLYGVSVMLLRGRRLLFSLGAPFRHGVKIAP 463
Db 380 -----EDLOGATVGOEPLTSVVRGGLVGYTVAMTSRSSRPLLAAGAPRYQHVGQVLLF 434
Qy 464 QL-KKQCAVRAVQASLOQEQIGSYFGSELCLPDTDRDGTDTDLVLAAPMFLPQNKETGRV 522
Db 435 QAPAGGRWNTQKIEGTQIGSYFGSELCSVDLDQGEAELLIGAPLFFGEQ--RGRV 492
Qy 523 YVYLVGQOSSLTTLQTLQPEPPQD--ARFGPAMGALPDNLQDGFADVAVAPLEBEGHGAL 581
Db 493 FTY-QRRQSLFEMVSELQDGPVPLGRFGAAITALTIDINGDRULTDVAVAPLEB--QCAV 549
Qy 582 YLHGTOGVRPHPAQRIAAASMPHALSYFGRSDVDRDLDDGLDLDVAVGAQGAAILLS 641
Db 550 YIFNGKPGGLSPQSPQRIQGAQVFFGIRWFRGIHGVKIDGDRDLADVVGAEGRVVVLS 609
Qy 642 SRPIVHLTPSLEVTPOAISVVQRDC---RRRQGEAVCLTAAALCFQVTSRTPGRWDHQFY 697
Db 610 SRPVVDVTELSFSPPEIPVHEVECSYSAREEQKGVKLA--CFRIKPLTP-----QFQ 662
Qy 698 MR-----FTASLDWETAGARAAAFDGSQRSLRRLRLSVGNV-----TCBQLHFHV-L 744
Db 663 GRLLANLSYTLQDGHMRSRGLFPDGSHEL-----GNTSITPDKSCLOPFHPFPI 714
Qy 745 DTSYLRPVALTVTFALDNTTKPGVNLNBSPT-----SIQKLVPFSKD 788
Db 715 CIQDLSPINVSNLNFSL-----LEEEGTDRQKGRAMQPILRPSIHVTTKEIPFKN 766
Qy 789 CGPDNECVTDLVLVQVNMIDIRGSRKAPFVVRGGRKRVLYSTLTLENKENAYNTSLISFSR 848
Db 767 CGEDKKCEANLTL-----SPASGGLRLMSS--ASLAVETWLSNGEDATVWRLDLPFR 820
Qy 849 NLHLASLTP-QRESPIKVECAAPSAHARL-----CSVGHPVQFGAKVTFLLPEFSCS 901
Db 821 GLSPRKVEMLQPHSRMPSVCEELTEGSSLLTKLKCNVSPPIFKAGQEV----- 870
Qy 902 SLLSQVFGKLTASS--DSLERNGTL-----QENTACTSAYIOYEPHLLFSS--S 947
Db 871 --LQVMFNTLVNSWEDFVELNGVHCENENSSIQEDNSAATHIPVLPVNLTKQENS 928
Qy 948 TLHRYEVHPYGTLPVGPSPFKTTLRTNNASCIVQNLTEPPGPPVHPPEELQHTNRLNGSN 1007
Db 929 TLY-----ISFPFKGP-----KTQQVQHVYQVRIQ---PSAYDHNPTLEALVG-- 969
Qy 1008 TQCOVVRCHLQOLAKGT-FVSVGLLLRLVHNEFFRRAKFKS----- 1046
```

Db 970 ---VPRHSEDLITYTWSVQTDPLVTCBSEDLKRPSSSEARQPCLPQGVQRCPIVFRWEI 1025
 Qy 1047 -LTVVSTFELGTE--EGSVQLQTEA-----SRWSES-----LLEVVQTRPI 1084
 Db 1026 LIQVTGVFVSKETKASSTLSLCSLSVSFNSSKHFLYSGKASEAQVLKVDLHKEKEM 1085
 Qy 1085 LISWILIGSVLGGLLALLVFLCWLKGLGPPFAHKKIPDEEKREKLE 1131
 Db 1086 L-HYVVLG--IGGLVLLFLFLALYKVGFF-----KRLKKEWE 1122

RESULT 8
 RWHUB
 cell surface glycoprotein CD11b precursor [validated] - human
 N;Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Ma
 eukocyte integrin alpha chain; neutrophil adherence receptor alphaM chain
 C;Species: Homo sapiens (man)
 C;Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text change 09-Jul-2004
 C;Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; I52567
 R;Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.
 J. Biol. Chem. 263, 12403-12411, 1988
 A;Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD
 B.
 A;Reference number: A31108; MUID:88315033; PMID:2457584
 A;Accession: A31108
 A;Molecule type: mRNA
 A;Residues: 1-1153 <COR>
 A;Cross-references: UNIPROT:P11215; GB:J03925; NID:G187284; PIDN:AAA59544.1; PID:G307148
 A;Note: part of this sequence was confirmed by protein sequencing
 R;Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.
 J. Cell Biol. 106, 2153-2158, 1988
 A;Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M
 A;Reference number: A28915; MUID:88257215; PMID:2454931
 A;Accession: A28915
 A;Molecule type: mRNA
 A;Residues: 1-499, 501-965, 'P', 967-1153 <ARN>
 A;Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:G186935; PIDN:AAA594
 A;Note: the authors translated the codon TAC for residue 1129 as Thr
 R;Shelley, C.S.; Arnaout, M.A.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991
 A;Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg
 A;Reference number: A41600; MUID:92073318; PMID:1683702
 A;Accession: A41600
 A;Molecule type: DNA
 A;Residues: 1-9 <SHE>
 A;Cross-references: GB:M76724; NID:G180018; PIDN:AAA58410.1; PID:G553215
 R;Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.
 Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988
 A;Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesi
 A;Reference number: A94193; MUID:88190151; PMID:2833753
 A;Accession: A30892
 A;Molecule type: mRNA
 A;Residues: 917-1042 <AR2>
 A;Cross-references: GB:M18044
 R;Hicksrein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.
 Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989
 A;Title: cDNA sequence for the alphaM subunit of the human neutrophil adherence receptor
 A;Reference number: A32218; MUID:89098893; PMID:2563162
 A;Accession: A32218
 A;Molecule type: mRNA
 A;Residues: 9-1153 <HIC>
 A;Cross-references: GB:J04145; NID:G189068; PIDN:AAA59903.1; PID:G386975
 A;Note: part of this sequence was confirmed by protein sequencing
 R;Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.
 J. Immunol. 150, 480-490, 1993
 A;Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-in
 n during evolution.
 A;Reference number: A46526; MUID:93123748; PMID:8419480
 A;Accession: A46526
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-499,501-1153 <FLE>

A;Cross-references: GB:S52227; NID:G263047; PIDN:AAB24821.1; PID:G263049
 A;Note: the last three bases of intron 13, CAG, are included in some but not all mature
 A;Note: sequence extracted from NCBI backbone (NCBIP:121963)
 R;Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.
 Biochim. Biophys. Acta 874, 368-371, 1986
 A;Title: N-terminal sequence of human leukocyte glycoprotein Mol.: conservation across sp
 A;Reference number: A90664; MUID:87076671; PMID:3539202
 A;Accession: A26091
 A;Molecule type: protein
 A;Residues: 17-31 <PIE>
 A;Experimental source: granulocytes
 R;Pahl, H.L.; Rogmarin, A.G.; Tenen, D.G.
 Blood 79, 865-870, 1992
 A;Title: Characterization of the myeloid-specific CD11b promoter.
 A;Reference number: I52567; MUID:9214986; PMID:1346576
 A;Accession: I52567
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-9 <RES>
 A;Cross-references: GB:M84477; NID:G180184; PIDN:AAA51960.1; PID:G553219
 C;Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1
 C;Genetics:
 A;Gene: GDB:ITGAM; CR3A
 A;Cross-references: GDB:120599; OMIM:120980
 A;Map position: 16p11.2-16p11.2
 A;Note: promoter contains a GATA motif and two Sp1 consensus binding sites
 C;Superfamily: cell surface glycoprotein CD1b; von Willebrand factor type A repeat homo
 C;Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag
 F;1-16/Domain: signal sequence #status predicted <SIG>
 F;17-1153/Product: cell surface glycoprotein CD1b #status experimental <MAT>
 F;17-1108/Domain: extracellular #status predicted <EXT>
 F;148-318/Domain: von Willebrand factor type A repeat homology <VWA>
 F;465-473/Region: calcium/magnesium binding #status predicted
 F;530-538/Region: calcium/magnesium binding #status predicted
 F;593-601/Region: calcium/magnesium binding #status predicted
 F;1109-1134/Domain: transmembrane #status predicted <TM>
 F;1135-1153/Domain: intracellular #status predicted <INT>
 F;86,240,331,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding
 Query Match 18.4%; Score 1091; DB 1; Length 1153;
 Best Local Similarity 29.6%; Pred. No. 1.5e-72;
 Matches 359; Conservative 202; Mismatches 482; Indels 170; Gaps 47;
 Qy 11 LPLVFLTG--LCSPFNLDHHPRLFPGPPEAFSGYSLVGHVGGQRMVLCAPWDGSGD 68
 Db 3 LRVLLLTALTLCGFGNLDTENAMTFQENARG-FGQSVVLQGSR---VVGAPQEIWAAN 58
 Qy 69 RRGDVYRCPVGAHNAFCAGHGLGDYQLGNSSHP-----AVNMHLGSLLETGDDGG 120
 Db 59 QRSLVQC-----DYSTG-SCEPIRLQVPVEAVNMSLGLSLAATTSPQ 101
 Qy 121 FMACAPLWSRACSSVSFSGICARVDASFPQGS LAPTAQR-CPTY-MDVVIVLDGNSI 178
 Db 102 LLACGPTVHQTCSNTVVGKLCFLFGSNLRQPKFPFALRGCPQEDSIAFLIDGSGSI 161
 Qy 179 YPNSVQTFRLRVGKLFIDP--EQIO-----VGLVOYSGSPVHWSLGFPRTKEEVVRA 231
 Db 162 IPHD-----FRFM--KEFVSTVMEQLKKSKTLFSLMQYSSEEFRIHFKFQNNPRSL 214
 Qy 232 AKNLSREGRQETKTAQAIWACTGEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALKA 291
 Db 215 VKPITQLLGR-THTATGIRKVKVRELFTNITGARKNPKILVITDGEKF-GDPLGYEDVI 272
 Qy 292 CEAGR--VTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNVTDSSAALTIDVA 349
 Db 273 PEADREGVIRYIGV-GDAFRSEKS-----RQELNTIASKPRDRHVFQVNNFEALKTQNQ 327
 Qy 350 LGDRIFGLEGSHAENESSFGLEMSIQGFSTHRLKQGLFGMVGAYDGGSVLMLEGGHRL 409
 Db 328 LREKIFAIEGTQTGSSSSSFEHMSQEGFSAAITNSGPLLSTLVGSDYWGAGVFLYTSREKS 387
 Qy 410 FPPMALEDEFPFPAQNHAAYLVGYSVSSMLLRCGRRLFLSGAPFRHGRKVIAQLKKDG 469

A;Title: Aberrant upregulation of a novel integrin alpha subunit gene at 3p21.3 in small
A;Reference number: 158409; MUID:94119603; PMID:8290272
A;Accession: J58409
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1035 <RES>
A;Cross-references: UNIPROT:Q13797; GB:D25303; NID:9464180; PIDN:BA04984.1; PID:G533327
R;Palmer, B.L.; Ruegg, C.; Ferrando, R.; Pytel, R.; Sheppard, D.
J. Cell Biol. 123, 1289-1297, 1993
A;Title: Sequence and tissue distribution of the integrin alpha9 subunit, a novel partne
A;Reference number: A49459; MUID:94064789; PMID:8245132
A;Accession: A49459
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 30-1035 <PAL>
A;Cross-references: GB:I24158
C;Superfamily: integrin alpha-4 chain
C;Keywords: glycoprotein; metal binding; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>

Query Match 12.0%; Score 714.5; DB 2; Length 1035;
Best Local Similarity 22.4%; Pred. No. 1.5e-44;
Matches 280; Conservative 170; Mismatches 432; Indels 369; Gaps 43;

Qy 9 LFLPLVELTGLCSFNLDEHHPLFPQPPAEFGYSVLQHVGGQRMVLCAPWD----G 64
Db 16 LLLALVAGITPAGAYNLDPQRPVHFQGPADSFYVLEHFDNTRVWLVGAPKADSKYS 75
Qy 65 PSGDRRGDVPCVGGAHNAP-----CAKHLGDYQLGNS-SHPANMHLQMSLL-E 114
Db 76 PSVKSPGAVFKRV---HTNDRCTELDMARGKRGTSCKTCREBDDWGVSLARQ 132
Qy 115 TDGSGFMACAPLWSRAC--GSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVIVL 172
Db 133 PKADGRVLACARWKNIVYEADHILPHGFCVLIIPSNLQAKG----- 173
Qy 173 DGSNIYPMSEVQTFRLRLVGLKLPIDPEIQVGLVQGESPVHWSLGDFTKSEVVRAA 232
Db 174 -----RT----- 175
Qy 233 KNLRSREGRETQATIAWVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKAC 292
Db 176 -----LIPCYEYKKY-----GEEHG-----SC 194
Qy 293 EAGRVTRYGIAVLGHYLRQRDPSPFLREIRTIASDPDERFFNVNTDEAALTDIVDALGD 352
Db 195 QAG-----IA----- 199
Qy 353 RIFLGSGHAENSSFGLEMSQIGFSTRHLKDGILFGMVGAYDNGGVSVMLEGGHRLPPP 412
Db 200 -----GFFTBEEL---VVMGAPGSFYWAGTIKVLNLTDNTY-- 231
Qy 413 RMALEDEPPPALQNHAAVLGYSVSSMLLRGRRLL-FLSGAPFRHRHGKVIQFOL-KKQGA 470
Db 232 -LKLNDV--IMNRYTYLVGAVTAGHFSHSTTDVVGGAQDQIGKVIYIFRADRSQT 288
Qy 471 VRVAQSLQGEIGYSGSELCPDTRDGTGTDVLLVAAPMFLGPQNKETGRVYVLYGQQ 530
Db 289 LKIFQAGSKWGSYFGSSLCVLDLNGDGLSD-LLVGAPMF--SEIRDEGQVTVINRGN 345
Qy 531 SLLTLQGTLPPEPPQDARFGFAMGALPDINQDGADVAAGAPLEDGHOGALYLYHGTQSG 590
Db 346 GALBEQALATGDGAYNAHFGEISLASLDLDNDGFPDVAIGAPKEDDFAGAYVYIHGDAGG 405
Qy 591 VRPHPAQRIAAAMPHALSYGRSVDGRDLDDGLVDVAVGA--QGAAILSSRPVHL 648
Db 406 IPVQYSMKLSGQKINPVLRMFQSGISGGIDMDNGYDPVTVGAFMSDVSLLRLARPVITV 465
Qy 649 TPLSEVTPQATISVQVQDCRRRQGEAVCLTAALCF-----QVTSRTP 689
Db 466 DVSI-FLPGSINITAPQCHDQQPQVNCVLTTCFSFGKHVPEEIGLNVLMADVAKKEK 524
Qy 690 GRWDHQFYMRFTASLDEWTAGARAAFDGSGQRLLSPRLRLSVGNVTCQLHFIHV-LDTS 748

Db 525 GQMPRVYFVLLGSTMGOVT-----EKQLTYMEETCHRVVAHVRRVQD 568
Qy 749 YLRPVALVTVFAL-----DNTTKP-GPVL--NEGSPTSIQKLVPFSCDGPNECVT 797
Db 569 VISPIVFEAAYSLEHVTGEEERELPELTPVLRAWKKGOKIAQKQTVFERNCRSE-DCAA 627
Qy 798 DLVLQVNDIRG-SRKAPFVVRGRRKVLVSTLLENKENAYNTLSLIIISRNHLASLT 856
Db 628 DLQLOGLKLLSSMDEKTYLALGAVKNISLNISINLGDADYANFVNSRELFNNMW 687
Qy 857 PORESPIKVECAAPSAHARLCSVGHVPVQTAGKVTFLLEFEFSCSSLSQVFGKL-TASS 915
Db 688 QKEE--MGISCLELSEDFKCSVGFPFMRSKSYEFSVIFDTHLSLEEEVLSFIVTAQS 745
Qy 916 DSLRNGTLOENTAQTSAYIQE-----PHLLFSSES----- 947
Db 746 GNTSESLDNTLVLMVPLMHEVDTSITGIMSPTSFVGYESVDAAANFIQLDDLECHPOP 805
Qy 948 ---TLHRYEVHPYGTLP-VGPGPEFKTTLTNNA-----SCIVQNLTEPP 988
Db 806 INITLQVYNTGP-STLPGSSVSISFPNRLSSGGAEHFVQEMVVGQKNGCSFQKNPTPC 864
Qy 989 GPPVHPELOHT-----NRLNGSNTQCVVRCHLGQAKGTEVSVGLRLVHN 1036
Db 865 IIFQEQENIFHTIFAPFTKGRKVLDCCKEKGISCLTAHCFSAKAKESRTIDIYMLNT 924
Qy 1037 E-----PFRRAKFK---SLTVVSTFELGTEGSLVQLTEASRWSSESLLEVQTR 1082
Db 925 EILKDDSSSVIQFMSRAKVKVDPALRVVEIAHGNEPVTV--VFEA-----LHNLEPR 975
Qy 1083 PILISLWILGVLGGLLLALLVFLCWKLGF--AHKKIPEEKREKLE 1131
Db 976 GYVVG-WIAISLVGLILFLLAVLLWKGFFRRYKEITEAEKRNENE 1025

RESULT 13
JC7294
alpha integrin - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004
A;Accession: JC7294
R;Susan, J.M.; Just, M.L.; Lennarz, W.J.
Biochem. Biophys. Res. Commun. 272, 929-935, 2000
A;Title: Cloning and characterization of alphaP integrin in embryos of the sea urchin St
A;Reference number: JC7294
A;Contents: Embryo
A;Accession: JC7294
A;Molecule type: mRNA
A;Residues: 1-1054 <SUS>
A;Cross-references: UNIPROT:Q9U6S1; GB:AAD55724
C;Genetics:
A;Gene: sualaphap
C;Superfamily: integrin alpha-2b chain
C;Keywords: calcium binding; embryo; glycolysis; glycoprotein; heterodimer; transmembran

Query Match 11.8%; Score 700; DB 2; Length 1054;
Best Local Similarity 22.7%; Pred. No. 1.9e-43;
Matches 286; Conservative 182; Mismatches 412; Indels 378; Gaps 45;

Qy 10 FLPLVFLITGL-----CSPFNLDHHPRLFPQPPAEFGYSVLQHVGGQRMVLYG---AP 61
Db 4 FLLLSFCTLLVLLDSTVAFNFDLRAPYKFDGPGQSLFGFSVAQHRDQNTDVLWLGAPAP 63
Qy 62 WDGPSSGRDRGDVRCYVGG-AHNAPCAK---GHLGDYQLGNSHSHPAVNMHGLMSLLETDG 117
Db 64 TTQPGVTNGAAYKCPVTPIPSGSCFCEQVPTDTTGTNEVLDKS-----NQWFGATLASSGP 119
Qy 118 DGGFMACAP--LWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVIVLDGS 175
Db 120 DGRILACAPRLVWLQT-----STISPT----- 141
Qy 176 NSIYPMSEVQTFRLRLVGLKLPIDPEIQVGLVQGESPVHWSLGDFTKSEVVRAAKNL 235

Db 142 ----- 141
Qy 236 SRREGRTKTAQAIWVACTBGFSGHGRPEAARLLVVVTDGSHDGEELPAALKACEAG 295
Db 142 -----DKERPT----- 148
Qy 296 RVTRYGIAVLGHVLRQRDPSSFLREIIRTASDPERFFNFVTDAAALTDIVDALGRIF 355
Db 149 -----GTCFVGH-----SDFTNFVNYSPCQSTDRD-----LF 175
Qy 356 GLBGSABNESSGLESQIGFSTHRLKDG--LLFGMVGYADWGSVLMLEGGHRLPPPR 413
Db 176 GFD-----KITHCOAGFSAQIPSDNSTLMVGAPGSY-----YLQG--QIPAQ 216
Qy 414 MA-LED-----EPPALQNAHAAYLGYSVSSMLLRG--GRRFLSGAPRFRH--RGKVIAPQL 465
Db 217 LSTLSVNSNTEQNAVFDN--SVRGYSALGDGDFNGDGLDYVVGTPGESLRLGLAIFD- 273
Qy 466 KKDGAVRVAQSLQGEQIGSYFGSELCLPDTDRDGTDLVLLVAAPMFL-----GPONKETGR 521
Db 274 --QSLVEIITPVVGEQIVSYFGYSVASVDVNGDGLDD--LLVGAPMFTNREPATEKWEAGR 330
Qy 522 VYVYL-----VGQSSLLTQGTLPQPEPPQDARFGFANGALPDNLQDGFADVAVGAPLE- 574
Db 331 VYVYLNADHSLGAPOMLTGKKI-----RARFGFPITSIGDSNQDGFNDVAIGAPYDG 383
Qy 575 DGHGALVLYHGTGSGVRPHPAQRIAAASMPHA--LSYFGRSVDGRLLDGDGLDVDAVGA 633
Db 384 EDNSGVYIYTHGSABGURLTESQVLTPELGFSDITTFGFSVGQDQMDQNDPDLVVG 443
Qy 634 QG--AAILLSSRPVHLTPSLVTPQAISSVVQRDCR--RRQGEAVCLTAALCFQVT--SRTP 689
Db 444 ESADAALVTRPVVLLAEALTIEPIGILNDKNKYELPDGTMTVTSFVAMACFTYTGHNLP 503
Qy 690 GRWDHQYPMFTASLDWBTAGARAAPDGSQORLSPPRLRLSVGNVTCEQLHFHVLDT--SD 748
Db 504 ARIGISYTLTVDDSI-----TSGRALLEVNLSQVTKNRNLVDGVMKFCPLRAYVYNTIQD 560
Qy 749 YLRPVALTVTFALDNTT-----KGPVNLNBSPTSIOKLVPFSKDCGPDNECVTDLVLQV 803
Db 561 KLTPVADVLQYELTDESILPYELLPIINKEAVSSQTKVSIQNNC--VNNICPIEIGITV 619
Qy 804 NMDIRGRKAPFVVRGRRKRVVSTTLNENKAYNTSLIIFSRNLHLASLTQPRESPI 863
Db 620 TPNL-----PNVIGQAQELTLVVSINNREGDAFQSTLAVYYPEGLQYVRLERRANWDF 673
Qy 864 KVECAAPSAHARL--CSVGHPVFTGAKVTFLEPFSCSLLSQVFK-----LTASS 915
Db 674 SVTCTEDSALRMITCTDGNPLVG-----KYNLEFGLTLSTL--QVSGDKDNIEFYLVA 726
Qy 916 DSLERNGTLOENTAQTSAIYQEPHLLPSSSES----- 947
Db 727 ENNEPDNLONNELNVTAIVADATLKLSSASYPEIVTVRVPEDNIVPEPFTKNASEADI 786
Qy 948 ---TLHRYEHPYG-----TL-----P 961
Db 787 GMEVHVLYEVRNTGSSNAEVTNLNIRWPKDENGDFYLLGIMTDSGVTQISQGANP 846
Qy 962 VGPQPEFTTLRTNNAACIVQNLTEPPGPPVHPPELQHTNRLNGSNTOCCQVVRCHLQOLA 1021
Db 847 LGVKLEASTKEQLSNSTQVSGRRKREG--YAEALQAQBPICFTPESCVLINCTIDEIK 904
Qy 1022 KGTBVSGLLRVHNEPFRAPKSLTVVSTFELGTESGV-----LQLTEASRW 1071
Db 905 ATKSVVRIILGRFWERTFOKAVSAVPVQVTLASTATATVRSIPYNIPLPMEFTDSTKA 964
Qy 1072 SESL--LEWVQTRPILISLWILIGSVLGLLLALLVFLCMLKGLGFAFKKIPPEEKREE 1128
Db 965 STLITABELVLPVVSIAWIIIVSVLGGIILILLIILGLMKCGFFERKKPGEDQKEYE 1022

A41131
lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse
N;Alternate names: integrin alpha-4
C;Species: Mus musculus (house mouse)
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 09-Jul-2004
C;Accession: A41131; S16742
R;Neuhaus, H.; Hu, M.C.T.; Hemler, M.E.; Takada, Y.; Holzmann, B.; Weissman, I.L.
J. Cell Biol. 115, 1149-1158, 1991
A;Title: Cloning and expression of cDNAs for the alpha subunit of the murine lymphocyte-
A;Reference number: A41131; MUID:92064645; PMID:1840602
A;Accession: A41131
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1039 <NEU>
A;Cross-references: UNIPROT:Q00651; EMBL:X53176; MID:951484; PIDN:CAA37316.1; PID:951485
C;Superfamily: integrin alpha-4 chain
C;Keywords: cytoskeleton; transmembrane protein
Query Match 11.2%; Score 666.5; DB 2; Length 1039;
Best Local Similarity 22.4%; Pred. No. 5.7e-41;
Matches 279; Conservative 184; Mismatches 382; Indels 403; Gaps 58;
Qy 23 FNLDDEHHRLFPQPPPEABEFGYSLVQHVGQGRMLVGP---W-DGPGDRRGDVRCPV 78
Db 41 YNLDPENALLYQPGSTLFGYSVYLHSHGSKRWLIIVGAPTASLWSNASVVPNGAIYRCGI 100
Qy 79 -----GGAHNAPCAKGLHDYQLGNSHSPAVNMHGLMSILLETGDG--GGFMAC 124
Db 101 RKNPNQTCBQLQSGSGSPGCGKTCLEERD-----NQLGVTLRSQPGNGSIVTC 151
Qy 125 APLWSRACGSSVSFSSGICARVDASFPQGS LAPTAQRCPTYMDVIVLDGNSIYPMSEV 184
Db 152 GHRW-----KNIF-----YMKSDNKLPT----- 169
Qy 185 OTFLRLVGLKFLDPRQIQVGLVOYGESPVHESLGDPRITKEVVRVAAKNLSRREGRETK 244
Db 170 -----GICYMP-----SDLRTE-----LSKR----- 186
Qy 245 TAQAIMVACTGEGFSQSHGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGI 304
Db 187 -----MAPCYKDYTRFGEN-----PASCQAG----- 208
Qy 305 LGHYLRQRDPSPFLREIIRTASDPDER-----FFNVTDEAALTDIVDALGRIFGLE 358
Db 209 -----ISSFYQDLIVMGAPGSSVYMTGVFVYNTI-----TNQYKAFVDR----- 248
Qy 359 GSHABNESSFGLBMSQIGFSTHRLKDGILLFGMVGYADWGSVLMLEGGHRLFPFMALED 418
Db 249 -----QNVKFG----- 255
Qy 419 EPPALQNAHAAYLGYSVSSMLLRGRRRL--FLSGAPRFRHGRKVIAFQKKDGAVRVAQSL 477
Db 256 -----SYLGYSVGAGHFRSPHTTEVVGAPQHEIGIKAYIIFI--DENELNIVYM 304
Qy 478 QGEQIGSYFGSELCLPDTDRDGTDLVLLVAAPMFLGPQN--KETGRVYVYLVQQS--LL 533
Db 305 KGKGLSGSYFGASCAVDNLADGFS--LLVGAPW---QSTIREGRVFEVYINGMGAMV 359
Qy 534 TLOGTLQPEPPQDARFGFANGALPDNLQDGFADVAVGAPLEDHQGLYLYHGTQSGVRP 593
Db 360 EMBERVLSGDKYAAAFGESEIANGLDNDGFDIAIGAPQEDDLRGAVIYNGRVDGISS 419
Qy 594 HPQRITAAASMPHALSYFGRSVDGRLLDGDGLDVDAVGA--QGAAILSSRPVHLTPS 651
Db 420 TYSQRTGEGQISLRLMFGQSIGQIDANNGYVDVAVGAFQSDSAVLLTRPVVIVEAS 479
Qy 652 LEVTPQAISSVQVDRRRRQGEAVCLTAALCFQVTSR--TPGRWDHQYPMFTASLDE---- 706
Db 480 LS-HPESVNRNTEKDCETENGLPSVCMHLLTLCFSYKGEVPG-----YIVLFVNVSLDVHRKA 534
Qy 707 -----WTAGARAAAFDGSQORLSPPRLRLSVGNVTCEQLH--FHVLDTSVLRPVALT 756
Db 535 ESPSRFYFFSNGTSDVITGS-----IRVSSGBKC-RTHQAFMRKQVRDILTPIHVE 585

Qy 757 VTFALDN--TTKPG-----PVLNKGSPS-IQKLVPFSKDCGPDNECVTVLQLVNM 805
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 586 ATYHLGHVITKRNTTEFPPLQPTLOKKKEGWIRKMNFARCAVEN-CSADLQVSXKV 644
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 806 D-IRGRKAPVVRGGRRKVLVSTLTNRKENAYNTSLIIFSRNLHLSLTLPQRSPIK 864
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 645 GFLKPENYKTYLAVGSMTIMLVNSLFNAGDDAYETTLNVOLPTGLYFIKILDLBKQIN 704
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 865 VECAAPSAAHRL-CSVGHPPVFQGA-K-VTFLLEFFPSCSLLSQVFGKLTAS-SDSLER 920
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 705 CEVTSSGIKVLACSLGYIVDRLSRIDISFLDV-----SSLRAHEDLSISVHASCEN 759
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 921 NGTL--QENTAOATSAYIQVEPHLLFSESSTLHRYEHP----YGTLPVGPGPEFKTTLR 973
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 760 EGEDLDQRDNRVTLTIPLRYEVML-----TVHGL-VNPTSFVYGSSEENE-PETCWAEK 811
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 974 TNNASCIQV-NLTEPPGPPVH-----PEELQHTNRLNGSN-T-QC----- 1010
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 812 LNLTFHVINTGISMAPNVSKIMVPSNFLPQQDKLFNVLDVQTTCQCHFKHYRECTFA 871
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 1011 -----QVVR-----CHLGOLAKGTEVSGVGLLRVLVHNE 1037
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 872 QOKGIAGTLTDIVKFSLTKDKRLLYCKMAQOHCIDFLCNFGKMGESKEAS-----VHIQ 925
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 1038 FFRRAKPKSLTVST--FELGT-----EGSVLQLTEASRWSESLLE-VVQTRP----- 1083
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 926 LEGRPSILEMDETSLSLKFEIKATAFPEHPKPVIELNKDENVAHFLEGLHHQRPKRFTI 985
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 1084 ILISLWLTIGSVLGGLLLLALLVFC-LWKLGPF-AHKKIPEBEKREE 1128
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 986 IITITISLLG-----LIIVLLISCVMWKAGFFRKQYKSIOLENRRD 1027
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 15
S06046
Integrin alpha-4 chain precursor - human
N;Alternate names: CD49d; very late antigen-4 alpha chain; VLA-4 alpha chain
C;Species: Homo sapiens (man)
C;Date: 01-Dec-1989 #sequence revision 01-Dec-1989 #text_change 09-Jul-2004
C;Accession: S06046; A39355; D28018
R;Takada, Y.; Ellices, M.J.; Crouse, C.; Hemler, M.E.
EMBO J. 8, 1361-1368, 1989
A;Title: The primary structure of the alpha(4) subunit of VLA-4: homology to other integrins
A;Reference number: S06046; MUID:89356603; PMID:2788572
A;Accession: S06046
A;Molecule type: mRNA
A;Residues: 1-1038 <TA>
A;Cross-references: UNIPROT:P13612; GB:X16983; EMBL:X15356; NID:g33945; PIDN:CAA34852.1; PRC;Rosen, G.D.; Birkenmeier, T.M.; Dean, D.C.
Proc. Natl. Acad. Sci. U.S.A. 88, 4094-4098, 1991
A;Title: Characterization of the alpha-4 integrin gene promoter.
A;Reference number: A39355; MUID:91239513; PMID:2034655
A;Accession: A39355
A;Molecule type: DNA
A;Residues: 1-93 <ROS>
A;Cross-references: GB:M62841
R;Takada, Y.; Strominger, J.L.; Hemler, M.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 3239-3243, 1987
A;Title: The very late antigen family of heterodimers is part of a superfamily of molecules
A;Reference number: A94151; MUID:87204112; PMID:3033641
A;Accession: D28018
A;Molecule type: protein
A;Residues: 40-50,'E',52-53 <TA2>
C;Genetics:
A;Gene: GDB:ITGA4; CD49D
A;Cross-references: GDB:I128032; OMIM:192975
A;Map position: 2q31-2q32
C;Superfamily: integrin alpha-4 chain
C;Keywords: cell adhesion; cytoskeleton; duplication; glycoprotein; heterodimer; transmembrane
F;1-39/Domain: signal sequence #status predicted <SIG>
F;40-1038/Product: integrin VLA-4 alpha-4 chain #status predicted <MAT>

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 167.54 Seconds
(without alignments)
3459.921 Million cell updates/sec

Title: US-09-647-544-4

Perfect score: 5912

Sequence: 1 MEKPFVTHFLPLVFLTGLC.....GFFAHKKIPBEKKKLEQ 1132

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.*

1: uniprot_sprot.*

2: uniprot_crembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5904.5	99.5	1167	1 ITAG HUMAN	O75578 homo sapien
2	2344.5	99.5	1189	1 ITAH HUMAN	Q9ukx5 homo sapien
3	2333.5	99.3	1188	1 ITAH MOUSE	P61622 mus musculus
4	2333.5	99.3	1188	2 QTQCC3	Q7tcq3 mus musculus
5	1884	31.8	1180	1 ITAL RAT	P18614 rattus norv
6	1881.5	31.7	1151	1 ITAL HUMAN	P56199 homo sapien
7	1827	30.8	1171	2 O42034	O42094 gallus gall
8	1734.5	29.2	1170	1 ITA2_BOVIN	P53710 bos taurus
9	1733.5	29.2	1181	1 ITA2_HUMAN	P17301 homo sapien
10	1712.5	28.9	1178	1 ITA2_MOUSE	Q62469 mus musculus
11	1712	28.9	1178	2 Q6P1C7	Q6p1c7 mus musculus
12	1403	23.7	823	2 Q8WY18	Q8wy18 homo sapien
13	1398.5	23.4	823	2 Q8CE94	Q8ce94 mus musculus
14	1135.5	19.1	1161	1 ITAD RAT	Q8eyr7 rattus norv
15	1126	19.0	1162	1 ITAD HUMAN	Q13349 homo sapien
16	1115.5	18.8	1170	1 ITAL HUMAN	P20701 homo sapien
17	1105.5	18.6	1163	1 ITAL_MOUSE	P24063 mus musculus
18	1104.5	18.6	1161	2 Q9WTV4	Q9wtv4 mus musculus
19	1097	18.5	1160	2 Q9R200	Q9r200 mus musculus
20	1090	18.4	1188	2 Q8KAS4	Q6kas4 mus musculus
21	1084.5	18.3	1152	1 ITAM HUMAN	P11215 homo sapien
22	1077.5	18.2	288	2 Q8BM12	Q8bm12 mus musculus
23	1073.5	18.1	1169	1 ITAX_MOUSE	Q9gmh4 mus musculus
24	1064	17.9	1165	1 ITAL_BOVIN	P61625 bos taurus
25	1062.5	17.9	1163	1 ITAX_HUMAN	P20702 homo sapien
26	1037	17.5	1166	2 Q6TYB8	Q6tyb8 bos taurus
27	1025	17.3	1153	1 ITAM_MOUSE	P05555 mus musculus
28	1013	17.1	1151	2 Q9J130	Q9ji30 rattus norv
29	994.5	16.8	1086	2 Q96HB1	Q96hb1 homo sapien
30	972.5	16.4	780	2 Q06Z71	Q06z71 xenopus lae
31	972	16.4	1196	2 Q98TF1	Q98tf1 cyprinus ca

32	956.5	16.1	1167	1 ITAE_MOUSE	O60677 mus musculus
33	956.5	16.1	1167	2 O88340	O88340 rattus norv
34	947.5	16.0	1167	2 O88341	O88341 rattus norv
35	937	15.8	1187	2 Q98TF0	P38570 cyprinus ca
36	928	15.6	1179	1 ITAE_HUMAN	P38570 homo sapien
37	893	15.1	920	2 Q28984	Q28984 sus scrofa
38	891.5	15.0	1038	2 Q8BS01	O8be01 mus musculus
39	883	14.9	895	2 Q9WUF8	Q9wuf8 mus sp. itg
40	873.5	14.7	927	2 Q8HZV0	Q8hzv0 bos taurus
41	792	13.4	1160	2 Q8MKF4	Q8mkf4 felis silve
42	780.5	13.2	1332	2 Q9BPQ8	Q9bpq8 halocynthia
43	754.5	12.7	1036	2 Q91YD5	Q91yd5 mus musculus
44	714.5	12.0	1035	1 ITA9_HUMAN	Q13797 homo sapien
45	704.5	11.9	1033	2 Q9BGU3	Q9bgu3 bos taurus

ALIGNMENTS

RESULT 1

ID	ITAG HUMAN	STANDARD;	PRT;	1167 AA.
AC	O75578; Q9UH28;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	05-JUN-2004 (Rel. 44, Last annotation update)			
DE	Integrin alpha-10 precursor.			
GN	Name=ITGA10;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Articular chondrocytes;			
RX	MEDLINE=98352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;			
RA	Camper L., Hellman U., Lundgren-Aakerlund E.,			
RT	"Isolation, cloning, and sequence analysis of the integrin subunit			
RT	alpha10, a beta1-associated collagen binding integrin expressed on			
RT	chondrocytes.";			
RL	J. Biol. Chem. 273:20383-20389(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Endothelial cells, and Heart;			
RX	MEDLINE=20169197; PubMed=10702680;			
RA	Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,			
RT	Wang S.-X., Langley R., Krissansen G.W.;			
RT	"The integrin alpha10 subunit: expression pattern, partial gene			
RT	structure, and chromosomal localization.";			
RL	Cytogenet. Cell Genet. 87:238-244(1999).			
CC	-!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.			
CC	-!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10			
CC	associates with beta-1.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: Widely expressed with highest expression in			
CC	muscle and heart. Found in articular cartilage.			
CC	-!- DOMAIN: The integrin I-domain (insert) is a VWF domain. Integrins			
CC	with I-domains do not undergo protease cleavage.			
CC	-!- SIMILARITY: Belongs to the integrin alpha chain family.			
CC	-!- SIMILARITY: Contains 7 FG-GAP repeats.			
CC	-!- SIMILARITY: Contains 1 VWF domain.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; AF074015; AAC31952.1; --
EMBL; AF112345; AAF21944.1; --
EMBL; AF172723; AAF61638.1; --

DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Integrin alpha-11 precursor.
 GN Name=ITGALL;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal heart, and Osteoblast;
 RX MEDLINE=99417678; PubMed=10486209; DOI=10.1006/geno.1999.5909;
 RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
 RA Wang S.-X., Morris C.M., Kriesansen G.W.;
 RT "Cloning, sequence analysis, and chromosomal localization of the novel
 RT human integrin alpha11 subunit (ITGALL).";
 RL Genomics 60:179-187(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal muscle, and Uterus;
 RX MEDLINE=99395147; PubMed=10464311; DOI=10.1074/jbc.274.36.25735;
 RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;
 RT "cDNA Cloning and Chromosomal Localization of Human alpha(11)
 RT Integrin. A collagen-binding, i domain-containing, beta(1)-associated
 RT integrin alpha-chain present in muscle tissues.";
 RL J. Biol. Chem. 274:25735-25742(1999).
 RN [3]
 RP SEQUENCE OF 954-1188 FROM N.A.
 RC TISSUE=Fibroblast;
 RA Andreu N., Estivill X., Escarceller M., Sunoy L.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen.
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11
 CC associates with beta-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: According Ref.2 highest levels in uterus and
 CC heart, intermediate levels in skeletal muscle and intermediate to
 CC low levels in pancreas, kidney and placenta. According to Ref.1
 CC also found in brain, colon, lung, small intestine, stomach,
 CC testis, salivary glands, thyroid glands and prostate. Very low
 CC levels in peripheral blood lymphocytes, fetal brain and fetal
 CC liver.
 CC -1- DEVELOPMENTAL STAGE: Strongly up-regulated in differentiating
 CC fetal muscle cells (in vitro).
 CC -1- DOMAIN: The integrin i-domain (insert) is a VWFA domain. Integrins
 CC with i-domains do not undergo protease cleavage.
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -1- SIMILARITY: Contains 1 VWFA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF109681; AA01258.1; -;
 CC EMBL; AF137378; AA051919.2; -;
 CC EMBL; AL359064; CAB94392.1; -;
 CC HSSP; P18614; IMHP.
 CC Genew; HGNC:6136; ITGALL.
 CC MIN; 604789; -;
 CC GO; GO:0008305; C:integrin complex; TAS.
 CC GO; GO:0005518; F:collagen binding; TAS.
 CC GO; GO:0007160; P:cell-matrix adhesion; TAS.
 CC GO; GO:0007517; P:muscle development; TAS.
 CC InterPro; IPR000413; Integrin_alpha.
 CC InterPro; IPR002035; VWFA.
 CC Pfam; PF00092; VWFA; 1.
 CC PRINTS; PR01185; INTEGRINA.
 CC PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWFA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
 DR PROSITE; PS0234; VWFA; 1.
 KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium;
 KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.
 FT SIGNAL 1 22 Potential
 FT CHAIN 23 1189 Integrin alpha-11.
 FT DOMAIN 23 1142 Extracellular (Potential).
 FT TRANSMEM 1143 1165 Potential.
 FT DOMAIN 1166 1189 Cytoplasmic (Potential).
 FT REPEAT 38 94 FG-GAP 1.
 FT REPEAT 102 163 FG-GAP 2.
 FT DOMAIN 164 345 VWFA.
 FT REPEAT 359 420 FG-GAP 3.
 FT REPEAT 422 475 FG-GAP 4.
 FT REPEAT 477 537 FG-GAP 5.
 FT REPEAT 539 598 FG-GAP 6.
 FT REPEAT 601 653 FG-GAP 7.
 FT DOMAIN 1154 1162 Poly-Leu.
 FT DOMAIN 1174 1177 Poly-Arg.
 FT CA BIND 488 496 Potential.
 FT CA BIND 551 559 Potential.
 FT CA BIND 613 621 Potential.
 FT DISULFID 76 83 By similarity.
 FT DISULFID 121 139 Potential.
 FT DISULFID 129 159 Potential.
 FT DISULFID 659 668 By similarity.
 FT DISULFID 674 729 By similarity.
 FT DISULFID 781 787 By similarity.
 FT DISULFID 881 893 By similarity.
 FT CARBOHYD 82 82 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 95 95 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 291 291 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 331 331 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 358 358 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 449 449 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 462 462 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 528 528 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 642 642 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 694 694 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 857 857 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 894 894 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 973 973 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1032 1032 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1040 1040 N-linked (GlcNAc...) (Potential).
 FT VARIANT 433 433 V -> M.
 FT VARIANT /FTID=VAR_009889.
 FT VARIANT 524 524 R -> L.
 FT VARIANT /FTID=VAR_009890.
 FT VARIANT 972 972 L -> P.
 FT VARIANT /FTID=VAR_009891.
 FT VARIANT 1003 1003 I -> M.
 FT VARIANT /FTID=VAR_009892.
 FT VARIANT 1030 1030 Missing.
 FT VARIANT /FTID=VAR_009893.
 FT VARIANT 1094 1094 L -> V.
 FT VARIANT /FTID=VAR_009894.
 SQ SEQUENCE 1189 AA; 133609 MW; 60303C08A4ACD52 CRC64;
 Query Match 39.5%; Score 2344.5; DB 1; Length 1189;
 Best Local Similarity 42.2%; Pred. No. 6.6e-159;
 Matches 505; Conservative 204; Mismatches 407; Indels 81; Gaps 16;
 Qy 1 MELPFVTHLPLVFLTGLCSFNLDEHRLPFGPPAEFGYVLQHVGGQRMVLVGA 60
 Db 1 MDLPRLVVAWALSLLWPGFTDTFNMDTRKRPVLPGSTAFPGYTVQQHDSGNKWLVGA 60
 Qy 61 PWDGPGSGDRRGDYYRCPVGGAHNAPCAKGLHGYQLGNSGHPAVNMHLGMSLLETGDDGG 120
 Db 61 PLEITNGYQKTDGVYKCPV---IHGNCCTKLNLGRVTLNLSERKDNMRGLSLATNPXONS 117
 Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGQSLAPTAQRCTPYMDVIVLDGNSIYP 180

Db 118 FLACPLWSEHCSSYYTTCMSRVNSFRFSKTVAPALQRCQTYMDIVVLGDSNIYP 177
Qy 181 WSEVQTLRLRLVGLFDPEIQOVLVOYGESPVHMSLGDFTKEBVRVRAAKNLSREG 240
Db 178 WVEVQHFLINLLKFFYIGPOIQGVQYQVEDVHHEFLNDYRSVKDNDVVEAASHIBGG 237
Qy 241 RETTAQAIWVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 238 TETRTAFGIFFASEAQK--GGRKGAKVMIVITDGESHSDPLEKVIQOESRDNTRY 295
Qy 301 GIAVLHYLRQRDPSPFLREIRTIASDPDERPFNVYTDAAALDIYDALGDRIFGLEGS 360
Db 296 AVAVLGYNRRGINPETFLEIKYIASDPDKPFNVYTDAAALDIYDALGDRIFSLGEGT 355
Qy 361 HAENESFGLMSQIGSTHRLKXGILFGWGVAGDMGSGVLMLEGHRLPPRMALEDEF 420
Db 356 N-KNETSFGLEMSQTGFSSHVEDGVLLGAVGDMGAVLKETSAGKVIPLRESYLKBF 414
Qy 421 PPALQNHAAVLGYSSVSMLLRGRRFLSGAPRFRHKGKVIQFOLKKDGVAVVAQSLQGE 480
Db 415 PEELKNHGAIVTYTTSVSSRQGRVVVAGAPRNFHTGKVLFTMHNRSILTIHQMRGQ 474
Qy 481 QIGSYFSGELCPDLTDRTDGLVLLVAAPLPGPQNKETGRVYVYLVGQSSLLTFLQSTLQ 540
Db 475 QIGSYFSGEITSVDIDGQVTDVLLVGAPMYFN--EGREGRKVVYVEL--RQNRFYVNGTLK 532
Qy 541 PEPP-QDARGCFANGALPDNLQDGFADVAGCAPLEDHOGALYLYHGTQGVRRHPAORI 599
Db 533 DSHSYQNRGSSIASVDRLQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPQRI 592
Qy 600 AAASMPHALSYFGRSDVGRDLDDGLDLDVAVAGQAAILSSRPVHLTPSLSEVTPQAI 659
Db 593 TASELATGLQYFGCSIHQGLDNEGLDILAVGALGNVILMSRPVQINASLHFESKI 652
Qy 660 SVVQDRCRRGQEAUCLTAALCFQVTSRTPGWMDHQFYMFRTASLDWNTAGAAAFDGG 719
Db 653 NIFHRDCKSRDQATCLAAFLCFTPIFLAPHFQTTTIGIRYNATMDERRYTTPRAHLDEGG 712
Qy 720 QRLSPRLRLSVGNVTCQLHFLVDTSDYLRLPVALTTFALDNTTKPGPVNLGSGPSTI 779
Db 713 DRFTNRVAVLSSGGELCEINFHVLDYADYVKPFTFSVEYSLEDPDH--GPMLLDGGWPTTL 771
Qy 780 QKLVFPKDCPDNECVTDLVQLVNMDI-----RGSRK-----APFV 817
Db 772 RVSVFPWNGCNEDEHCVPDLVDARSDLPTAMEYQORVLRKPAQDCSAYTILSPDTTVPIL 831
Qy 818 RGGRRKVLVSTTLNRKENAYNTSLIFSRNLHLASLTPORESPIKVECAAPS--AHAR 875
Db 832 ESTQRVAVATLENRGENAYSTVLNISQSANLQFASLIQKEDSGSIECWNERRLQKQ 891
Qy 876 LCSVGHVPVQTGAKVTLLEPFSCSLLSVOFKLTASSDSLERNGTLOENTTAQTSAYI 935
Db 892 VCNVSYFFFRAKAVAPRLDPEFSKIFLHHLETELAAGSDNSRDKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSESTLHRYEVHPYGLP--VGPGEFKTTLRTN----- 975
Db 952 KYEADVLFTRSSLSHVEKLVNLSLERYDGIQPPFCIFRQNLGLPFIHGMKTIPI 1011
Qy 976 -----NASC-IVONLTTEPPGVPHPELQHTNRLNGSNNTQCQVVR 1014
Db 1012 ATRSGNRLKLRLDPLTDEVANTSCNIWGNSTERYPTPE--EDLRAPLQNLHNSDVSIN 1070
Qy 1015 CHLQOLAGTREVSVGLRLVHNEFRRAKPSLTVSTFELGTSEGSVLQTEASRWSES 1074
Db 1071 CNI--RLVNPQBNFHLGNLWRLSKALKYKSMKIMVNAALQROFHPFPFIFREDBPQRQI 1129
Qy 1075 LLEVVTQRPILISWLITIGSVLGGALLLALVFLCKLWGLFFPAHKIPEEEKREKLE 1131
Db 1130 VPEISKQEDWQVPIWIVGSTLGGULLLALLVLALMLKLGFRSAR----RRREFGLD 1182

ITAH MOUSE
ID ITAH MOUSE STANDARD; PRT; 1188 AA.
AC P61622;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-11 precursor.
GN Name=Itgall;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Wozny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahay J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen (By
similarity).
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11
associates with beta-1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
with I-domains do not undergo protease cleavage.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; BC058716; AAH58716.1; -;
CC MGD; MGI:2442114; Itgall.
CC PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
CC PROSITE; PS02034; VWFA; 1.
CC Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
CC Repeat; Signal; Transmembrane.
CC SIGNAL 1 22 Potential.
FT CHAIN 23 1188 Integrin alpha-11.
FT DOMAIN 23 1141 Extracellular (Potential).
FT TRANSMEM 1142 1164 Potential.
FT DOMAIN 1165 1188 Cytoplasmic (Potential).
FT REPEAT 38 94 FG-GAP 1.
FT REPEAT 102 163 FG-GAP 2.
FT REPEAT 164 345 VWFA.
FT DOMAIN 359 420 FG-GAP 3.
FT REPEAT 422 475 FG-GAP 4.
FT REPEAT 477 537 FG-GAP 5.
FT REPEAT 539 598 FG-GAP 6.

FT	REPEAT	601	653	FG-CAP 7.	
FT	CA_BIND	488	496	Potential.	
FT	CA_BIND	551	559	Potential.	
FT	CA_BIND	613	621	Potential.	
FT	DISULFID	76	83	By similarity.	
FT	DISULFID	121	139	Potential.	
FT	DISULFID	129	159	Potential.	
FT	DISULFID	659	668	By similarity.	
FT	DISULFID	774	789	By similarity.	
FT	DISULFID	781	787	By similarity.	
FT	DISULFID	881	893	By similarity.	
FT	CARBOHYD	82	82	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	95	95	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	291	291	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	331	331	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	358	358	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	449	449	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	462	462	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	528	528	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	642	642	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	694	694	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	857	857	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	894	894	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	973	973	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	1031	1031	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	1188	1188	N-linked (GlcNAc. .)	(Potential).
FT	SEQUENCE	1188	1188	AA; 133011 MW; 0B0313C90D65422E CRC64;	
Query Match					39,3%; Score 2333.5; DB 1; Length 1188;
Best Local Similarity					42.2%; Pred. No. 4e-158;
Matches					499; Conservative 200; Mismatches 398; Indels 85; Gaps 15;
Qy	18	GLCSFPNLDHPRLPFGPPEAEFGYSVLQVGGQRMVLPAPWDGPGSDRRGDVYRCP	77		
Db	18	GFTDTFMDTRNPRVIAGPSAARFGYTVQQHDSGKWLUVGAPMETNGHGTGDVYKCP	77		
Qy	78	VGGAHNAPCAKHGLDYLQGNSSHPVNMHLCMSLLETDDGGGFMACAPLWSRACGSSVF	137		
Db	78	VTQGN---CTKLNLRVTLNVSERKDNMRGLSLATNPKNDSFLACSLWSHCEGSSYY	134		
Qy	138	SSGICARVDASFOFGSLAPTAQCPCTVMDVVVLVDGNSIYPMSEVQTFRLRLVGLFI	197		
Db	135	TTGMSRANSNFRSKVPALQRCQTMIDIVLDGNSIYPMVEVQHFILNLIKFYI	194		
Qy	198	DPEIQVLQVQGESPVHWSLGRTEKEEVVRAAKNLSRREGRETKTAQAIWVACTEGF	257		
Db	195	GPQIQVGIQVQGEDAVHEFLNDYRSVKOVVEASHIEQRGGTETRTAFGIEFARSEAF	254		
Qy	258	SQSHGGRPEARLLVVVTVDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSS	317		
Db	255	QK--GGRKGAKVMIVITDGSHPDLEKVIQSEKDNVTRYAVAVLGYNNRGINPET	312		
Qy	318	FLRIRTIASDPDRFFENVTDALTDIVDALGRIFLGEGSHAENESSFGLMSQIGF	377		
Db	313	FLNEIKYIASDPDKFENVTDALTDIVDALGRIFLGEGTN-KNETSFGLMSQIGF	371		
Qy	378	STHRLKDGILFGMVGADWGSVLWLEGGHRLFPFRMALEDFPPALQNHAAIYGVSVSS	437		
Db	372	SSHVVEDGILLGAVGADWNGAVLKETSAGKVIPIHRESYLKEFPEELKNHAAIYGVTVTS	431		
Qy	438	MLLGGRRLLFLSGAPRFRHRKGVIAFOLKKDGVAVVAQSLQGEQIGSYFGSELCPFLDTR	497		
Db	432	VWSSRQGRVYVAGAPRFRNHTKVLFSMHNRSRLTIHQALRGEQIGSYFGSEITSVDVD	491		
Qy	498	DGTTDVLVAPMELGPONKETGRVYVYLVQCOSSLITLQTLQPEPP-QDARFGFAMCAL	556		
Db	492	DRVTDVLVAGPMYFS-EGRGRKVVYVNL-RQNRVYVNGTLKDSHYSQNAFSGSCIASV	549		
Qy	557	PDLNQDGFADVAVGAPLEDGHQALYLYHGTQSGVRPHPAQRIAAASPHALSIFYGRSVD	616		
Db	550	QDLNQDSYNDVVVAGAPLEDHSHRGAIYIPFGQTWILKKPMQRIYASELAPQLQHFQGSIH	609		

Qy	617	GRLLDGLDLDVAVGAAGAAIILSSRPVHLTPSLEVTPOAITSVVORDCRRRQOEAVCL	676		
Db	610	QQLDLEDGLDVLAVGALGNVAVLWAPVQIINASHLFESKINIFHKCKRNGRDATCL	669		
Qy	677	TAALCFQVTSRTPGRMWDHQFYMFRTASLDEWTAGARAAFDGSGQRLSPRLRLSVGNVTC	736		
Db	670	AAFCLFPIELAPHFQATVIGIRYNATMDERRYMPRAHLDEGGDQFTNRAVLLSSGQEH	729		
Qy	737	EQLHFHVLDSYLRPVVALVTTFALDNTTKPGPVNLEGSPTSQKLPFSDKDCGPDNECV	796		
Db	730	QRINFHVLDTADYVVPKPVAFVSVEYSLEDPDN-GPMLDNGWPTTLRLVSYPFWNGCNEDEHCV	788		
Qy	797	TDVLVQVMDIRGSRK-----APFVVRGGRKVLVSTTLENRK	834		
Db	789	PDVLVDARSDLPAMEYQVRPAQDCSSYLSFTDTTFTTERRRVAEATLENRG	848		
Qy	835	ENAYNTSLSIIFSRNLHLASLTPORESPIKVECAAPS--AHARLCVSGHPVFOFGAKVTF	892		
Db	849	ENAYSAVLNISQSENLFASLIQKDDSDNIECVNEBRLHKKVCNVSYPFRAKAKVAF	908		
Qy	893	LLPEFSCSLLSQVFGKLTASSDSLERNGLTQNTAQTSAITQYEPHLLFSSESTLHRY	952		
Db	909	RLDFEFSKSVFLHHLQIHLGAGSDSHEQDSTADNTALLRFLKYEADVLFTRSSLSHF	968		
Qy	953	EVHPYGTLPV--GPGPEFKTLRT-----IPEEEK	1125		
Db	969	EVKANSLESYDGIQPPFNCVQVQNLGFFPIHGVMKTIPTIATRGNRLMLLRDFTD	1028		
Qy	975	NNASC-IVONLTPEPPVHPBELOHTNRLNGSNTQCQVVRCHLGOLAKGTEVSVGLLR	1032		
Db	1029	QGNSTCNINWNSYEYRSTPTE-EDLSHAPORHNSNDVSIICNL-RLASQETSFYLVG	1086		
Qy	1033	LVNNEFPRAKFKSLTVVSTFELTEGSGVLQTEASRWSESLLEVQTRPILSLMILI	1092		
Db	1087	NLWLTSLKALKYRSKLITVNAALQRFSPFIFREEDPSQVTFEISKQEDWQVPIWII	1146		
Qy	1093	GSVLGGLLALLVFLWKLGFPAHKK-----IPEEEK	1125		
Db	1147	GSTLGGLLALLVIALWKLGFPAHKK-----IPEEEK	1188		
RESULT 4					
ID	Q7TQC3	PRELIMINARY;	PRT;	1188	AA.
AC	Q7TQC3;				
DT	01-OCT-2003	(TrEMBLrel. 25, Created)			
DT	01-OCT-2003	(TrEMBLrel. 25, Last sequence update)			
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)			
DE	All integrin.				
GN	Name=Itgall;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	PubMed=15183724; DOI=10.1016/j.ydbio.2004.03.006;				
RA	Popova S.N., Rodriguez-Sanchez B., Liden A., Betscholtz C.,				
RA	van den Bos T., Gullberg D.,				
FT	"The mesenchymal alpha1beta1 integrin attenuates PDGF-BB-stimulated				
RT	chemotaxis of embryonic fibroblasts on collagens.";				
RL	Dev. Biol. 270:427-442 (2004).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Johansson M., Popova S.N.;				
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.				
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).				
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.				
DR	EMBL; AY124460; AAM62130.1; -				
DR	HSP; P18614; 1CK4.				
DR	GO; GO:0006929; P:substrate-bound cell migration; IMP.				
DR	InterPro; IPR000413; Integrin_alpha.				
DR	InterPro; IPR002035; VWF_A.				

```
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; VWA; 1.
DR SMART; SM00191; Int alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS02034; VWA; 1.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 1188 AA; 133067 MW; 69B2BF20D75E4271 CRC64;

Query Match          39.3%; Score 2333.5; DB 2; Length 1188;
Best Local Similarity 42.2%; Pred. No. 4e-158;
Matches 499; Conservative 200; Mismatches 398; Indels 85; Gaps 15;

QY 18 GLCSPPNLDHHPRLPFGPPEAEGYSVLQVGGQRMMLVGAPMDGPSGDRRDGVYRCP 77
DB 18 GFTDFNMTRNPRVAGSAAFPFGTVQQHDISGKKWLVVAGPMETNGHQTKGDTYKCP 77
QY 78 VGGHNAFAPCAKGLHDYQLGNSHPA VNMHGLMSLLETGDDGGPMACAPLMSRACGSVF 137
DB 78 VTQGN---CTKLNIGRVTLGNVSRKDNMRMLGSLATNPKNDSPLACSLWSHECGSSYY 134
QY 138 SSGICARVDASFOQGS LAPTAQRCPTMYMDVIVLDGNSNIYPMSEVQTFRLRLVGLKFI 197
DB 135 TTGCSRVNSNFRSKTVAPALQRCQTYMDIVIVLDGNSNIYPMSEVQTFRLRLVGLKFI 194
QY 198 DPEQIQVGLVOYGSPPVHNSLGDFTKKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGF 257
DB 195 GPGQIQVGIQVQGEDA VHEPHLNDYRSKVDVVEAASHIEQRGGTETRTAFGIEPAREAF 254
QY 258 SQSHGGPEEARLLVVVTDGESHGDEELPAALKACEAGRVTRYGIAVLGHLYLRQRDPSS 317
DB 255 QK--GGRKAGKAKMIVITDGHSDSPDLKVIROSEKDNTRYAVAVLGYNNRGINPET 312
QY 318 FLRIRTIASDPDRFPFNVTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSOIGF 377
DB 313 FLNEIKTIASDPDKHFNFTVDEAALDKDIVDALGDRIFSLGNTN-KNETSFGLMSQTGP 371
QY 378 STHRLKDGILFGMVGAYDMGQSVLWLEGHRLFPFRMALEDEFPALQNAAYLGYSVSS 437
DB 372 SSHVVEDGILLGAVGAYDMGAVLKETSAGKVI PHRESYLKEPPEELKNHVAIYLGTVTS 431
QY 438 MLLRGGRLFLSGAPRFHKGKVIATFQKKGAVRVAQSLQGEIQISYFGSELCPDTR 497
DB 432 WSSRQGRVYVAGAPRNHTGKVLFLSMHNNRSITIHQALRGEIQISYFGSEITSVSDVND 491
QY 498 DGTDDLVAAPMPLGPQNKETGRVYVYVVGQSQSLTLTQGTLOPEPP-QDARFGFAMGAL 556
DB 492 DRVTDVLLVAGPMYFS-EGRGRKVVYNNL-RQRFVYNGTLKDSHYSQNAFQSCIASV 549
QY 557 PDLNQDGFADVAVGAPLEDGHQALYLYHGQSGVRPHPAQRIAAASHPHALSIFYGRSVD 616
DB 550 QDLNQDSYNDVVVGA PLDSDHRGAIYIFHGFTWILKKPMQORITASELAPGLQHFQCSIH 609
QY 617 GRLLDDGDLVDVAVGAQAAILLSRPVHLTSPLEVTPOAISVVDORCRRQEAACL 676
DB 610 GQLDNLNEDGLVDLAVGALGNAVLWAPVPVQINASLHFEPSKINIFHKDCKRNGRDATCL 669
QY 677 TAALCFQVTSRTPCRWDHQFVMTFASLDEWTACARAAFDGSGORLSRRLRLSVGNVTC 736
DB 670 AAFLCFPIFLAFHFQATVGIIRYNATWDERYMPRAHLDEGGQFTNRAVLLSSGQEH 729
QY 737 EQLHFHVLDTSDYLRPVALVTFALDNTTKPGPVINEGSPITSIOKLVPFSDKDCPDNECV 796
DB 730 QRINFHVLDTADYVKVAFSVEYSLEDPDN-GPMLDNGWPTTLRVSPVFWNMGNCDEHCV 788
QY 797 TDLVLQVNMIDRGRK-----APVVGGRKVLVSTLTLENRK 834
DB 789 PDLVLDARSDLPTAMEYQCQVLRGPAQCSSYTLSTFDTTVFIESTRRRVAVEATLENRG 848
QY 835 ENAYNTSLSIFSNLNLHSLTPQRESPKVECAAPS--AHARLCSVGHVPVFOGAKVTF 892
DB 849 ENAYSAVLNTSQSENLOFASLIQKDDSDNSIECVNEERLHKVCNVSYPPFRKAKAVP 908
QY 893 LLEFEFSCSLLSQVFGKLTASSDSLENGTLQENTAQTSAYIQYEPHLLFSSSESTLHRY 952
```

DR PDB; 1MHP; X-ray; A/B=169-360.
 DR RGB; 2923; Itgal.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 3.
 DR Pfam; PF00357; Integrin_alpha; 1.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS0234; VWA; 1.
 DR 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;
 KW Magnesium; Receptor; Repeat; Signal; Transmembrane.
 FT CHAIN 1 28
 FT DOMAIN 29 1180 Integrin alpha-1.
 FT TRANSMEM 1143 1165 Extracellular (Potential).
 FT DOMAIN 1166 1180 Potential.
 FT REPEAT 104 103 Cytoplasmic (Potential).
 FT REPEAT 175 388 FG-GAP 1.
 FT REPEAT 377 432 FG-GAP 2.
 FT REPEAT 433 484 FG-GAP 3.
 FT REPEAT 485 565 FG-GAP 4.
 FT REPEAT 567 626 FG-GAP 5.
 FT REPEAT 629 681 FG-GAP 6.
 FT CA_BIND 497 505 FG-GAP 7.
 FT CA_BIND 579 587 Potential.
 FT CA_BIND 641 649 Potential.
 FT SITE 1168 1172 GFFKR motif.
 FT DISULFID 82 92 By similarity.
 FT DISULFID 687 696 By similarity.
 FT DISULFID 702 755 By similarity.
 FT DISULFID 807 813 By similarity.
 FT DISULFID 877 885 By similarity.
 FT DISULFID 1029 1062 By similarity.
 FT DISULFID 1066 1073 By similarity.
 FT CARBOHYD 100 100 N-linked (GlcNAc...)
 FT CARBOHYD 105 105 N-linked (GlcNAc...)
 FT CARBOHYD 112 112 N-linked (GlcNAc...)
 FT CARBOHYD 217 217 N-linked (GlcNAc...)
 FT CARBOHYD 317 317 N-linked (GlcNAc...)
 FT CARBOHYD 341 341 N-linked (GlcNAc...)
 FT CARBOHYD 402 402 N-linked (GlcNAc...)
 FT CARBOHYD 418 418 N-linked (GlcNAc...)
 FT CARBOHYD 459 459 N-linked (GlcNAc...)
 FT CARBOHYD 531 531 N-linked (GlcNAc...)
 FT CARBOHYD 698 698 N-linked (GlcNAc...)
 FT CARBOHYD 747 747 N-linked (GlcNAc...)
 FT CARBOHYD 779 779 N-linked (GlcNAc...)
 FT CARBOHYD 820 820 N-linked (GlcNAc...)
 FT CARBOHYD 839 839 N-linked (GlcNAc...)
 FT CARBOHYD 882 882 N-linked (GlcNAc...)
 FT CARBOHYD 907 907 N-linked (GlcNAc...)
 FT CARBOHYD 938 938 N-linked (GlcNAc...)
 FT CARBOHYD 965 965 N-linked (GlcNAc...)
 FT CARBOHYD 973 973 N-linked (GlcNAc...)
 FT CARBOHYD 1007 1007 N-linked (GlcNAc...)
 FT CARBOHYD 1084 1084 N-linked (GlcNAc...)
 FT CARBOHYD 1103 1103 N-linked (GlcNAc...)
 FT CARBOHYD 1114 1114 N-linked (GlcNAc...)
 FT STRAND 172 178
 FT HELIX 186 197
 FT TURN 198 199
 FT TURN 204 205
 FT STRAND 208 214
 FT STRAND 218 222
 FT TURN 224 226
 FT TURN 230 238
 FT TURN 239 239
 FT HELIX 250 259

FT TURN 260 261
 FT HELIX 264 266
 FT TURN 267 267
 FT TURN 270 271
 FT STRAND 274 280
 FT TURN 285 286
 FT HELIX 287 289
 FT HELIX 290 299
 FT TURN 300 301
 FT STRAND 302 309
 FT HELIX 311 315
 FT TURN 316 317
 FT HELIX 321 328
 FT TURN 329 330
 FT HELIX 335 338
 FT STRAND 339 342
 FT HELIX 346 350
 FT HELIX 352 359
 SQ SEQUENCE 1180 AA; 130808 MW; 8E5DA2BE02362EE4 CRC64;
 Query Match 31.8%; Score 1884; DB 1; Length 1180;
 Best Local Similarity 36.2%; Pred. No. 7.6e-126;
 Matches 432; Conservative 219; Mismatches 432; Indels 112; Gaps 25;
 QY 13 LVFLTGLCSPLNLDHHPRLPPGPEAEFGYSLVHQVGGQRMWLVGAPWDPSPGSDREGD 72
 DB 19 LTVLGFVSNFVDVKNMSFSGPVEDMFGYVQOYENEGKMWLIGSLPLVGQPKARTGD 78
 QY 73 VYRCPVGGAHNAPCAKAGHLGDYQLGNSSHPAV-----NMHLGMSLLETGDDGGFMACAPL 127
 DB 79 VYKCPGGERAMPCKLDLP-----VNTSIPNVTIKENMTFG-STLVNPNNGFLACGPL 133
 QY 128 WSRACGSSVFSGICARVDASFPQGS LAPPAQCPTVMVIVLDGNSNIYPMSEVQTF 187
 DB 134 YAVRCGHLHTTGICSDVSPTFQVNSFAP-VQECSTQLDIVIVLDGNSNIYPMSEVIAF 192
 QY 188 LRLVGLKFLDPEQIOVLVOYGESPVHWSLGDFTKEEVVRAAKNLSRREGRETAKQ 247
 DB 193 LNDLLKRMKGPKOTQVGIQOYGENVTHEFNLNKYSSTEEVLVAANKIGRGGLOTMTAL 252
 QY 248 AIMVACTGFSQSHGGRPEARLLVVTGDESHDGEELPAALKACAGRVTRYGI AVLGH 307
 DB 253 GIDTARKEATEARGARGVKVMVIVTDESHDNYRLKQVIOCEDENIQRFSAI LGH 312
 QY 308 YLRQRDPSSFLREIRTIASDPDRFFNFVTVDEAALTDIVDALGDRIFGLGSHAES 367
 DB 313 YNRGNLSTKRFVEIKSIASEPTKHFNFVSDDELAVTIKALGERIFALEATADQSAAS 372
 QY 368 FGLEMSQIGFSTHRLKDGILFGMVGAYDWGSSVLWLSGGHRLFPFRALEDEFPALQNH 427
 DB 373 FEMMSQTGFSAHYSQDWMVLGAVYDWNGVTVMQKANKQVIPHNTTTFQTE--PAK 430
 QY 428 --AAYLGVSYSSMLLRGRRFLTSGAPFRHRGKVI AFOLKKDGAVRVAQSLQGEQIGSY 485
 DB 431 PLASYLGYTVNSATIPGD-VLYIAGQRYNHTGVVIYKM-EDGNINILQTLGGEQIGSY 488
 QY 486 FGSELCLDTRDGTDTDLVLAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTLP---- 541
 DB 489 FGSVLTITIDKDSYTDLLLVGAPMVGTEKEEGKVYVAV-NQTRFEYQMSLEPIRQT 547
 QY 542 -----EPQDARFGFANGALPDNLQDGFADVAVGAPLEDHGHQALYLYHG 586
 DB 548 CCSSLKDNCSCTKENKNEPCGARFGTAIAVKDLNVDGNDVVI GAPLEDHAGAVIYHG 607
 QY 587 TQSGVRPHPAQRIAAASMPHALSYFGRSVDGRDLDDGLVDVAVGAQAAILSSRPV 646
 DB 608 SGKTI REAYAQRI PSGGDGKTLKFFGSIHGEMLDNGDGLTDVTIGLGGALFWARDVA 667
 QY 647 HLTPSLEVTPQAI SVVQDRCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMTASLDE 706
 DB 668 VVKVTMNFEPKNVNIQKNCRVECKETVCINATWCFFHVKLSKEDSIYEADLQYRVTLDS 727
 QY 707 WTAGARAAFDCSGQRRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTFFALDNTTK 766

```
Db 728 LRQISRFSGTQERKIQR--NITVRESECIHRSFYMLDKHDFQDSVRVTLDLFD---NLTD 782
Qy 767 P--GPVLNKGSPSTQIKLVKPSKDCGPDNCEVTDLVQVNDIRGSRKAPVVRGGRKV 824
Db 783 PENGVLDDALPNSVHHIIPAKDCGKERCISDLTNVST----TEKSLIVKSHQDKP 838
Qy 825 LVSTTLNKRKENAYNTSLSIIFSRLHLASLTP-QRESPIKVECAAPSAHARLCSVGHVP 883
Db 839 NVSLTVNKGDSAYNTRTVQHSPLNLFSGIEEIQKDS-----CESNQNIICRVGYPF 891
Qy 884 FQCAKQVTFLEPFSSGSSLLSQVFGKLTASSDLSRNGTLQENTAQTSAYIQEPHLF 943
Db 892 LRAGETVFKIIFQNTSHLSNAIIHLSATSDEEPLSLNDNEVNSIPVKYEVGLQF 951
Qy 944 SSBSTLHRYVHPVGTLP-----VG-----PGRPFKTL----- 972
Db 952 YSSASEHHISVAANETIPEFINSTEDIGNEINVYTIKRGHPMPPELQJLSFPNLTAD 1011
Qy 973 -----RTNNAACIVQNLTEPPG-----PPVHPBELQHTNRLNGSNTQCQVVR 1015
Db 1012 GYPVLYPIGWSSSDNVNCRPSLEDPGINSKGKWTISKSEVLKRGTIQDCSSTCGVATI 1071
Qy 1016 HLGQLAKG-TEVSVGLRLVHNEFRRAKFKSLTVVSTFELGTGEGSVLQLTASRWSES 1074
Db 1072 TCSLLPSDLSQVNSL--LLWKPTFIRAHFSSLLTLRGELKSENS--LTLSNRRKREL 1128
Qy 1075 LLEVQVQ-RTLISLWLIQSLVGLGLLLALLVCLWKLGFPAHKKIPSEKREE 1128
Db 1129 AIQSKDGLQGRVPLWILLISAFAGLLLLMLLIALWLKIGF---KRPLKKKMEK 1180

RESULT 6
ITAL_HUMAN
ID _ITAL_HUMAN STANDARD; PRT; 1151 AA.
AC P56199;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
GN Name=ITGAL;
OS Homo sapiens (Human).
OC Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euteleostomi; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93155124; PubMed=8428973;
RX Brileschitz R., Epstein M.R., Marcantonio E.E.;
RT "Expression of native and truncated forms of the human integrin alpha
RT 1 subunit."
RL J. Biol. Chem. 268:2989-2996(1993).
CC -!- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and
CC collagen. It recognizes the proline-hydroxylated sequence G-P-P-G-
CC E-R in collagen.
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1
CC associates with beta-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: The integrin I-domain (insert) is a VWF domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- SIMILARITY: Contains 1 VWF domain.
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD49a entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd49a.htm".
DR PIR; A45226; A45226.
DR PDB; 1QC5; X-ray; A/B=140-331.
DR PDB; 1QCY; X-ray; A=141-333.
DR Genew; HGNC:6134; ITGA1.
DR MIM; 192968;
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0005518; F:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; NAS.
```

```
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00244; VWF_A; 1.
KW 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;
KW Magnesium; Receptor; Repeat; Transmembrane.
FT DOMAIN 1 1113 Extracellular (Potential).
FT TRANSMEM 1114 1136 Potential.
FT DOMAIN 1137 1151 Cytoplasmic (Potential).
FT REPEAT 16 75 FG-GAP 1.
FT REPEAT 76 75 FG-GAP 2.
FT DOMAIN 147 360 VWF_A.
FT REPEAT 349 404 FG-GAP 3.
FT REPEAT 405 457 FG-GAP 4.
FT REPEAT 459 520 FG-GAP 5.
FT REPEAT 540 599 FG-GAP 6.
FT REPEAT 602 654 FG-GAP 7.
FT CA_BIND 470 478 Potential.
FT CA_BIND 552 560 Potential.
FT CA_BIND 614 622 Potential.
FT SITE 1139 1142 GPFKR motif.
FT DISULFID 54 64 By similarity.
FT DISULFID 660 669 By similarity.
FT DISULFID 675 728 By similarity.
FT DISULFID 780 786 By similarity.
FT DISULFID 858 858 By similarity.
FT DISULFID 1002 1034 By similarity.
FT DISULFID 1037 1044 By similarity.
FT CARBOHYD 46 46 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 72 72 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 77 77 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 84 84 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 189 189 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 289 289 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 313 313 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 374 374 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 390 390 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 432 432 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 504 504 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 671 671 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 720 720 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 752 752 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 812 812 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 855 855 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 880 880 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 887 887 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 911 911 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 938 938 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 946 946 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 980 980 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1055 1055 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1074 1074 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1085 1085 N-linked (GlcNAc...) (Potential).
FT STRAND 143 150 N-linked (GlcNAc...) (Potential).
FT TURN 153 154
FT TURN 158 158
FT TURN 159 159
FT HELIX 170 171
FT TURN 174 174
FT STRAND 176 177
FT TURN 177 177
FT STRAND 179 186
FT STRAND 190 194
FT TURN 196 197
FT TURN 211 211
FT HELIX 222 222
FT HELIX 223 223
```

FT	TURN	232	234	713	QERKQVOR	--NITVRKSECTKHSFYMLDKHDFQDSVRITLDF---	NLTPDENGPPVLDSDSLP	767
FT	HELIX	236	238	777	TSIQKLVPSKDCGPDNECVTDVLQVNMDIRGSRKAPFVVVGRGRRKVLSTTLENRKEN		836	
FT	TURN	239	239	768	NSVHEYIPFAKDCGKNCISDLHLV----	ATTEKOLLIVRSQNDKFNVLTKNTKDS	823	
FT	STRAND	242	243	837	AYNTSLSIIPSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPVQTCAGKVTFLLEF		896	
FT	TURN	245	252	824	AYNTRIVHYSPLNVFSGI----	EAIQKDSG--ESNHNITCKVGYPLFRRGEMWTFKLIF	877	
FT	HELIX	257	258	897	EFSCSLLSQVFGKLTASSDLSRNGTLOENTAQTSAYIQVEPHLLFSSESSTLHRYEHP		956	
FT	HELIX	262	271	878	QNTSYLMENVTVIYLSATSDEEPPETLSNVVNVISIPVKEVGLQFVSSASEVHISIAA		937	
FT	TURN	272	273	957	YGTLP-----VG-----	PGBPFTTLRTNNASCIVQNLTEPPGPPVH	993	
FT	STRAND	274	281	938	NETVPEVINSTEDIGNEINIFYLIRKSGSPMPBLKLSIS-----	FPNMTSGYPVLY	990	
FT	HELIX	283	287	994	PEELQHTNRLN-----	GSNTQCQ-----	VVRCHLGQA	1021
FT	TURN	288	289	991	PTGLSSSENANCRPHIFEDPFSINSKGKMTTSTDHLKRGTLDCNTCKFATITCNLTS-S		1049	
FT	HELIX	307	310	1022	KGTEVSVGLLRVHNEFFRRAKFKSLTVVSTFELGTTEEGSVLQLTEARMSSESLEVVQT		1081	
FT	STRAND	311	314	1050	DISQVNVSL--ILWKPTFIKSYFSSNLNLTIRGELRSENAS-LVLSSSNQKRELAIQISKD		1106	
FT	HELIX	317	323	1082	RPILSLTWLIGSVLGGLLLLALLVFLCWLKLPFAHKKPIPEEKREE		1128	
FT	HELIX	324	330	1107	GLPGRVPLWVILLISAFAGLLLLMLLILALWKIGFF---	KRPLKKMBEK	1151	
Qy	SEQUENCE	1151 AA;	127837 MW;	6B3F3CIAABF52808 CRC64;				
Query Match								
Best Local Similarity 35.9%; Pred. No. 1.1e-125;								
Matches 426; Conservative 220; Mismatches 423; Indels 119; Gaps 23;								
Qy	23	FNLDHHPRLPPGPEAEFGYSVLQHVGGGQRMMLVGAPEWDPGSDRRGDVYRCPVCGAH	82					
Db	1	FNVDKNSMTFSGVEDMFVTVQOYENEEKVVLGSLVPGQPNRTGDKYKCPVGRGE	60					
Qy	83	NAPCAKHLGDYQIGNSHPAV-----NMHLGMSLLETDGDPGMACAPLMSRACGSSVF	137					
Db	61	SLPCVKLDLP---VNTSIPNVTEVKENMTFG-STLVTPNPGGFLACGPLYAVRCGHLHY	115					
Qy	138	SSGICARVDASFOGSLAPTAQCPYMDVIVVLDGNSIYPWSEVQTFRLRLVGLKFI	197					
Db	116	TTGICSDVSPFTQVNVSIAP-VQECSTQLDVIIVLDGNSIYPWDSVTAFLNLDLLKMDI	174					
Qy	198	DPEIQIQLVQYSGSPVHWSLQDFRTKEEVVRAAKNLSRREGRETTKTAQAIMVACTEGF	257					
Db	175	GPKTQTVGIQVGENVTHEFNLNKYSSTEEVLVAAKIVQGGQWTWALTGDTARKEAF	234					
Qy	258	SQSHGRPEARLLVVTVDGSHDGEELPAALKAACEAGRVTTRYGIAVLGHYLRQRDPSS	317					
Db	235	TEARGARRGVKVMVITVDGSHDNHRLKKVIQDCEDENIQFSAIILGYSYNRGNLSTEK	294					
Qy	318	FLRIRTIASDPDRFPFNVTDAAALTDIVDALGRIFGLEGSHAENESSFGLMSQIGF	377					
Db	295	FVEBKSIASEPTKHPNFVSDALVTVIKTLGERIPALEATADQSAASFEMEMSQTF	354					
Qy	378	STHRLKDGILFGMYGVDWGGSVLWLEGGHRLFPFRMALEDEFPPALQNHAAYLGYSVSS	437					
Db	355	SAHYSQDWMLGAVGVDWGTVMQKASQIIIPNRTFNVESTKKEPLASYLYGVNS	414					
Qy	438	MLLRGRRRLFLSGAPRFRHKGKVIAPQLKDGAVRVAQSLQGEIQISYFGSELCPDLTDR	497					
Db	415	ATASSGDVLYIAGQPRYNHTQVLIYRM-EDGNIKILQTLGSEIQISYFGSILTTDIDK	473					
Qy	498	DGTTDVLVAAPMFLGPQNKETGRVYVVLVQSQSLLTLLQTLQTP-----	541					
Db	474	DSNTDILLVAGPMYMGTEKEEQKVYVAL-NQTRFEYQMSLEPIKQTCSSRQHSCTT	532					
Qy	542	---EPQDARFEGFAMGALPDINODGFADVAVGAPLEDGHQCALVYHGTQSGVRPHPAQR	598					
Db	533	ENKNEPCARGFTAAVAVKDLNLDGNDIVIGAPLEDHGHGAVYIHGSGTKIRKFAQR	592					
Qy	599	IAAASMPHALSYGRSDGRDLDDLDVVDVAVGAQGAAILLSRPVHLTPSLEVTQA	658					
Db	593	IPSGDGGKTLKFFQSGIHGEMDLNGDGLTDVTIGLGAALFWSRDVAVKVTVNFEFNK	652					
Qy	659	ISVVQDRCRRGQAEVCLTAALCFQVTSRTFGRWDHOFYMRFTASLDEWTAGAAAFDGS	718					
Db	653	VNIQKNCHEMGKETVCINATVCEVFKLSKEDTYEADLQYRVTLDSLRLQISRSFSGT	712					
Qy	719	GQRLSPRLRLSVGNVCEQLHFFHVLDTSDYLRPVALTVTALONTYKTP--GPVLNCSGP	776					

[illegible]

RX MEDLINE=89308879; PubMed=2545729; DOI=10.1083/jcb.109.1.397;
RA Takada Y., Hemler M.E.;
RT "The primary structure of the VLA-2/collagen receptor alpha 2 subunit
RT (platelet GPIa): homology to other integrins and the presence of a
RT possible collagen-binding domain.";
RL J. Cell Biol. 109:397-407(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Rieder M.J., Arnel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;
RT "SeattleSNPs, NHLBI HUG66682 program for genomic applications, UW-
RT FHCRC, Seattle, WA (URL: <http://pga.gs.washington.edu>).";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.
RX MEDLINE=98019223; PubMed=9353312; DOI=10.1074/jbc.272.45.28512;
RA Emaley J., King S.L., Bergelson J.M., Liddington R.C.;
RT "Crystal structure of the I domain from integrin alpha2beta1.";
RL J. Biol. Chem. 272:28512-28517(1997).
RN [4]
RP VARIANT GLU-534.
RX MEDLINE=94043762; PubMed=7901236;
RA Santoso S., Kalb R., Walka M., Kiefel V., Mueller-Eckhardt C.,
RA Newman P.J.;
RT "The human platelet alloantigens Br(a) and Brb are associated with a
RT single amino acid polymorphism on glycoprotein Ia (integrin subunit
RT alpha 2).";
RL J. Clin. Invest. 92:2427-2432(1993).
RN [5]
RP VARIANT GLU-534.
RX MEDLINE=20206009; PubMed=10744142;
RA Kroll H., Gardemann A., Fechter A., Haberbosch W., Santoso S.;
RT "The impact of the glycoprotein Ia collagen receptor subunit A1648G
RT gene polymorphism on coronary artery disease and acute myocardial
RT infarction.";
RL Thromb. Haemost. 83:392-396(2000).
CC -1- FUNCTION: Integrin alpha-2/beta-1 is a receptor for laminin,
CC collagen, collagen C-propeptides, fibronectin and E-cadherin. It
CC recognizes the proline-hydroxylated sequence G-P-G-E-R in
CC collagen. It is responsible for adhesion of platelets and other
CC cells to collagens, modulation of collagen and collagenase gene
CC expression, force generation and organization of newly synthesized
CC extracellular matrix.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
CC associates with beta-1. Interacts with HPSS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: The integrin I-domain (Insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- POLYMORPHISM: Position 534 is associated with platelet-specific
CC alloantigen HPA-5 (Br). HPA-5A/Br(a) has Lys-534 and HPA-5B/Br(b)
CC has Glu-534. HPA-5B is involved in neonatal alloimmune
CC thrombocytopenia (NAIT or NATP). The Lys-534-Glu polymorphism may
CC play a role in coronary artery disease (CAD).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VWFA domain.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD49b entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd49b.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X17033; CAA34894.1; --
CC EMBL; AF512556; AAM34795.1; --
CC PIR; A33998; A33998
CC PDB; 1A0X; X-ray; A/B=166-368.
CC PDB; 1DZ1; X-ray; A=171-355.
CC Genew; HGNC:6137; ITGA2.

DR MIM; 192974; --
DR GO; 0008305; C: integrin complex; TAS.
DR GO; 0005886; C: plasma membrane; TAS.
DR GO; 0005518; P: collagen binding; TAS.
DR GO; 0007596; P: blood coagulation; TAS.
DR GO; 0007160; P: cell-matrix adhesion; TAS.
DR GO; 0009887; P: organogenesis; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP_2.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWFA; 1.
DR PRINTS; PR01185; INTEGRIN.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWFA_1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PSS0234; VWFA; 1.
DR 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;
KW Glycoprotein; Integrin; Magnesium; Platelet; Polymorphism; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 29 Integrin alpha-2.
FT CHAIN 30 1181 Extracellular (Potential).
FT DOMAIN 1133 1154 Potential.
FT TRANSMEM 1155 1161 Cytoplasmic (Potential).
FT DOMAIN 1155 1161 Interaction with HPSS.
FT REPEAT 45 103 FG-GAP 1.
FT REPEAT 104 188 FG-GAP 2.
FT DOMAIN 188 378 VWFA.
FT REPEAT 379 433 FG-GAP 3.
FT REPEAT 434 486 FG-GAP 4.
FT REPEAT 488 549 FG-GAP 5.
FT REPEAT 551 610 FG-GAP 6.
FT REPEAT 615 667 FG-GAP 7.
FT CA BIND 499 507 Potential.
FT CA BIND 563 571 Potential.
FT CA BIND 627 635 Potential.
FT SITE 1157 1161 GPFKR motif.
FT DISULFID 83 92 By similarity.
FT DISULFID 680 737 By similarity.
FT DISULFID 789 795 By similarity.
FT DISULFID 865 876 By similarity.
FT DISULFID 1019 1050 By similarity.
FT DISULFID 1055 1060 By similarity.
FT CARBOHYD 105 105 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 112 112 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 343 343 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 432 432 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 460 460 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 475 475 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 699 699 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1057 1057 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1074 1074 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1081 1081 N-linked (GlcNAc...) (Potential).
FT VARIANT 534 534 K -> E (in alloantigen HPA-5B;
FT dbSNP:1801106).
FT FTId=VAR_003977.
FT TURN 170 171
FT STRAND 173 180
FT TURN 183 184
FT HELIX 188 199
FT TURN 200 201
FT STRAND 204 204
FT TURN 206 207
FT STRAND 209 216
FT STRAND 220 224
FT TURN 226 228
FT HELIX 232 240
FT TURN 241 241
FT HELIX 252 262
FT TURN 262 264
FT HELIX 266 268

CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -|- DOMAIN: The integrin I-domain (insert) is a VWF domain. Integrins
 CC with I-domains do not undergo protease cleavage.
 CC -|- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -|- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -|- SIMILARITY: Contains 1 VWF domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z29987; CA82877.1; -;
 DR EMBL; X75427; CA53178.1; -;
 DR PIR; S44142; S44142.
 DR HSP; P17301; IAOX.
 DR MGD; MGI:96600; Itga2.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR02035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 2.
 DR Pfam; PF00357; Integrin_alpha; 1.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS0234; VWF; 1.
 KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;
 KW Receptor; Repeat; Signal; Transmembrane.
 FT SIGNAL 1 26
 FT CHAIN 27 1178
 FT DOMAIN 27 1129 Integrin_alpha-2.
 FT TRANSMEM 1130 1151 Extracellular (Potential).
 FT DOMAIN 1152 1178 Potential.
 FT REPEAT 42 100 Cytoplasmic (Potential).
 FT REPEAT 101 7 FG-GAP 1.
 FT REPEAT 185 7 FG-GAP 2.
 FT REPEAT 376 430 VWF.
 FT REPEAT 431 483 FG-GAP 3.
 FT REPEAT 485 546 FG-GAP 4.
 FT REPEAT 548 607 FG-GAP 5.
 FT REPEAT 612 664 FG-GAP 6.
 FT REPEAT 624 682 FG-GAP 7.
 FT CA_BIND 496 504 Potential.
 FT CA_BIND 560 568 Potential.
 FT CA_BIND 624 632 Potential.
 FT SITE 480 482 Cell attachment site (Potential).
 FT SITE 1154 1158 GPPKR motif.
 FT DISULFID 80 89 By similarity.
 FT DISULFID 677 734 By similarity.
 FT DISULFID 786 792 By similarity.
 FT DISULFID 862 873 By similarity.
 FT DISULFID 1016 1047 By similarity.
 FT DISULFID 1052 1057 By similarity.
 FT CARBOHYD 102 102 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 109 109 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 429 429 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 457 457 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 472 472 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 696 696 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1054 1054 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1071 1071 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1078 1078 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;
 Query Match 28.9%; Score 1712.5; DB 1; Length 1178;
 Best Local Similarity 34.7%; Pred. No. 1.6e-113;
 Matches 412; Conservative 217; Mismatches 457; Indels 103; Gaps 32;
 9 LFPLVLFLT-GL---CSPFNLDHHPRLFPGPPEAFGYSVLQHVGGGQRMWLVGAPWDGP 65

Db 10 LLQLLMLVQGIINCLAYNVGLFGAKIFSGPSSEQFGYSVQQLTNQGNWLLVGSWSGF 69
 Qy 66 SGRRGDVTRCPVGGAAHNAKCAKGH- GDYQLGNSHPAVNMHLGMSLLETDCDGGPMAC 124
 Db 70 PENRMDGVKCPV- DLPATCEKLNQNGASISNVTETKTNMSLGULTLRNPGTGGGLTC 128
 Qy 125 APIWSRACSSSVSSGICARVDASFPQGS LAPTAQRCPYMDVIVLVDGNSIYPMSEV 184
 Db 129 GPLMAHQCNGYATGICSDVSPDFQLASFSPAVQACSLVDVVVVCDNSNIYPWEAV 188
 Qy 185 QTFRLRLVGLKFTDPQIQVGLVQGESPVHWSLGDFTKBEVVAAKNLSRREGRETK 244
 Db 189 KNFLVFVTGLDTPGKKTQVALIQYANEPRIIFNLNDFETKEDVMQVQATSETROGGDLTN 248
 Qy 245 TAOAIMVACTEGFSQSHGGRPEARLLVVVTCGESHGDELPALAKACEAGRVTYGIIV 304
 Db 249 TFRATFARDYAYQTSGGRPATKVMVVVTDGESHGSKLKTIVIOQNDDELIRFGIAV 308
 Qy 305 LGHYLRQRDPSSFLREIRTIASDPDERFFNVTDEAALTDIVDALQDRIFLGGSHAEN 364
 Db 309 LGYLNALDITKNLKEIKAIASPTPTERYFFNVADAEALKEKAGTLGEQIFSEGT-VQG 367
 Qy 365 ESSFGLEMSQIGPSTHRL--KQILFGMVGAYDWGSGVLMLEGHR--LPFPRMALEDEF 420
 Db 368 GDNFQEMAQVGFSAADYAPQNDILMLGAVGAFDWSGLV-QETSHKVPVIPP---KQAF 421
 Qy 421 PPALQ--NHAAYLGYSVSSMLLGGRRFLSGAPRFRHGRKVIAPQLKKGAVRVAOSLQ 478
 Db 422 DQVLQDRNHSSFLGYSVAALSTEDGVH-FVAGAPRANYTQGVLYSVNKGNTVVIQSHR 480
 Qy 479 GEQIGYFSGELCPDTRDGTDDVLLVAAFMPLGPQNKETGRVYVYLVQOQSLLTLQGT 538
 Db 481 GDQIGYFSGVLCSDVDKDTITDVLVGPATYMDLKEEGKVYLTITKGINLHQFTL 540
 Qy 539 LQPEPQDARFGPAMGALPDNLQDGPADVAAGPLEDGHOGALYLYHGTGSGVRPHPAQR 598
 Db 541 EGPEGTGNARFGSAIAALSINNMGDFNDVIGSVENENSGAVIYVNGHQGTIRTKYSQK 600
 Qy 599 IAAA--SMPHALSVFGSRVDRLDGDGLVDVAVGAQGAAILLSRPIVHLTPSLVPTP 656
 Db 601 ILGNGAFRRHLQFFGRSLDGYGLNGDSITDVSIGALQVQLMSQSIADVAIEALFTP 660
 Qy 657 QAISVVQDRCRRRQGEAVCLTAALCFQVTSRTSGRMDHQFYMFTASLDLDTWAGAAAFD 716
 Db 661 DKITLTKDAK-----ITLKLCPRAEPRPAGQ--NNQVAILFNMTLDADGHSSRVTSR 711
 Qy 717 GSGQRLSPRLR--LSVGNV-TCEQLHFHVLDTSDYLRPVALVTVPALDNTWPKG--PVL 771
 Db 712 GVFPRESEFLQNMVNVQKCEHHISIQKPSDVVNPLDLRDVDSLLEN---PGTSPAL 768
 Qy 772 NEGSPTSIQKLVPSKDCGPDNECVTDVLQVNMDIRGSKAPVVRGRRKVLVSTLE 831
 Db 769 EAYSETVKVFSIPFYKECGSDGICISDLIDV-QQLPAIQTSQFIVSNQNKRTFFSVILK 827
 Qy 832 NRXENAYNTSLSIIFSRNHLASLTQRESPI---KVECAAPSAHARL-CSVGHPVPTG 887
 Db 828 NRGESAYTVVLAEPSENFFASF--NPVDGTEVTCVGSQSKSVTCVGVPAKSE 883
 Qy 888 AKVTFLEFPFGSSLLSVFGKLTASSDSLENGTLQENTAOQTSAYIQVEPHLLPSSS 947
 Db 884 QQVTFITNFDNLQNLQNAAINFOAFSESQETNKA--DNSVSLTIPLLDAELHLTRST 941
 Qy 948 TLHRYEVHPYGTLP-----VGPQ-----PEEK----- 969
 Db 942 NINFEISSDENAPSVKSVEDIGPKFIPSLKVTAGSAPVSMALVTTHIIOYTKKKNPLL 1001
 Qy 970 --TTLRTNNA---SCI--VQNLTEP---PGPPVHPELQHTNRLNGSNTQCVVRCCHLQ 1019
 Db 1002 YLTGIQTDQAGDISCTAEINPLKLPHTAPSVSKPENFRHTKELDCRTTSCNITCWLD 1061
 Qy 1020 LAKGTEVSVGLLKVHNEFFRRAKPKSLTVSTVFELGTGEGSVLQLTGAEARWSESLLLEV 1079

Db 1062 LHMKAEBFVNTTRVNRTPAASFTQVLTAAAEIDTHNPQ-LFVIEENAVTIPLMIMK 1120

Qy 1080 QTRPILISLWLGSLVGLLILLALLVFLCWLKLGFF--AHKXI---PEE 1123

Db 1121 PTEKAEVPTGVIGSIITAGIILLALLAMTAGLWKLKGFRRQYKQMGQNPDE 1169

RESULT 11

Q6P1C7 PRELIMINARY; PRT; 1178 AA.

AC Q6P1C7; STRAIN=C57BL/6; TISSUE=Eye;

DT 05-JUL-2004 (TREMBLrel. 27, Created)

DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)

DE Integrin alpha 2.

GN NamesItga2;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Eye; DOI=10.1073/pnas.242603899;

RC MEDLINE=22388257; PubMed=12477932;

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derce J.G.

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Haieh F.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Casavant T.L., Scheetz T.E.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Eye;

RA Strauberg R.

RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).

CC -1- SIMILARITY: Belongs to the Integrin alpha chain family.

DR EMBL; BC065139; AAH65139.1; -.

DR HSSP; P17301; 1A0X.

DR GO; GO:0005615; C:extracellular space; TAS.

DR GO; GO:0016021; C:integral to membrane; TAS.

DR InterPro; IPR00413; Integrin_alpha.

DR InterPro; IPR02035; VWF_A.

DR Pfam; PF01839; FG-GAP; 2.

DR Pfam; PF00357; Integrin_alpha; 1.

DR Pfam; PF00092; VWA; 1.

DR PRINTS; PR01185; INTEGRIN.

DR PRINTS; PR00453; VWFADOMAIN.

DR SMART; SM00191; Int alpha; 5.

DR SMART; SM00327; VWA; 1.

DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.

DR PROSITE; PS02034; VWFA; 1.

KW Cell adhesion; Integrin; Transmembrane.

SQ SEQUENCE 1178 AA; 128954 MW; 62FAEA820242A9B6 CRC64;

Query Match 28.9%; Score 1712; DB 2; Length 1178;

Best Local Similarity 34.58; Pred. No. 1.7e-113;

Matches 405; Conservative 218; Mismatches 455; Indels 102; Gaps 31;

Qy 13 LVFLTGL--CSPENLDEHHPRLPFGPEAEFGYSVLQGVGGQRMWLVGAPWDGSPGDRR 70

Db 15 LMDVQGLNCLATNVLGPGAKIFSGSSSEQFGISVQQLTNPQGNWLVGSPGSPENRM 74

Qy 71 GDVYRCFVGGAHNAKCAKHL-GDYQLGNSHSPAVNMHLGMSLLETDDGCGFACAPLWS 129

Db 75 GDVYKCPV-DLPTATCEKLNQNSASISNVTEIKTNMGLGLTRNPGTGGLTCGFLWA 133

Qy 130 RACGSSVSSGICARVDASFPQGSGLAPTAQRCEPTYMDVIVLDGNSIYPWSEVQFLR 189

Db 134 HQCGNQYATGICSDVSPDFQFLTSPFAVQACPSLVDDVVVVCDESNIYPWEAVKQFLV 193

Qy 190 RLVGKLFIDPEQIQVGLVQGESPVHWSLGDRTKEEVVRAAKNLRRREGRETKTAQAI 249

Db 194 KFTVGLDIGPKTQVALIQIANEPRIIFNLNDFETKEDMVQATSETQHGGLDINTFRAI 253

Qy 250 MVACTEGFSQSHGGRPEARLLVVVTDGESHGDEELPAALKACEAGRVTRYGIYAVLGHYL 309

Db 254 EFARDVAYSQTSGRPGCATKVMVVVTDGESHGDSKLTQVIQCNDDDEILRFGIYAVLYLN 313

Qy 310 RRORDPSSFLREIRTTASDPDERFFNVDBAALTDIVDALGDRIFGLESHAENESSFG 369

Db 314 RNALDTKNLIKEIKATASTPTERYFFNVADAALKEKAGTLGEQIFSIETG-VQGGNFQ 372

Qy 370 LEMSOIGFSTHRL--KDGILFGVMGVADWCGSVLWLBGGHR--LFPFMALEDEFPALQ 425

Db 373 MENAQVGFSAADYAPQNDILMLGAVGAPDWSGTLV-QETSHKPVIFP-----KQAFDQVLQ 426

Qy 426 --NHAAYLVGYSVSMILRGRRLLFLSGAPFRHRGKVIAPQLKDGAVRVAQSQQGQIG 483

Db 427 DRNHSSFLGYSVAIAISTEDGVH-FVAGAPRANYTGQIVLVSVNKGQNVTVIQSHRGDQIG 485

Qy 484 SYFGSELCPIDTRDGTDTLLVAAMPFLGPQNKETGRVTVLVGQOSLTLTQOTLOPEP 543

Db 486 SYFGSVLCSVDVDKDTITDVLVCGAPTYMNDLKEGKGYLFTTTKILGHQHQLEGPEG 545

Qy 544 PQARFGFAGALPDNLNODGFADVAVGAPLEDHGQALYLYHGPOSVRHPAPQRIAAA- 602

Db 546 TGNARFGSAALASDINMGDFNDVGVSPVENENSGAVIYNGHQGTIRTKYSKILGSN 605

Qy 603 -SMPHALSYFGRSVDGRDLDDLDVAVGACAAIILSSRPVHLTPSLEVTPQAIISV 661

Db 606 GAFERHLQFFGRSLDGYGLNGDSITDVISGALGOVQLMSQSIADVAIEALFPDKITL 665

Qy 662 VQRDCRRRGQEAIVCLTAALCFQVTSRTPRPGRDHQFYMRFTASLDDEWTAGARAAFDGSGQR 721

Db 666 LNKDAK-----ITLKCFAEFEPAGQ-NNQVAILFNMTLDADGSHSRVTSRGVFRE 716

Qy 722 LSPRLR--LSVGNV-TCEQLHFHVLDTSDYLRPVALTVTFALDNTTKPG--PVLNKGSP 776

Db 717 NSERFLQKNMVNVEVQKSEHHISIQKPSDVANPLDLRVDISLEN---PGTSPALEAYSE 773

Qy 777 TSIQKLVFFSKDCGPDNECVTDLVQVNMIDIRSKAPFVVVGRGRRKVLVSTLTLENKEN 836

Db 774 TVKVFSIPFYKEGSDGICISDILLDV-QQLPAIQOSFIVSNQKSLTSLVILKNRES 832

Qy 837 AYNTSLSIIFSRNLHLASLTPQRESPI---KVECAAPSAAHARL-CSVGHVPVFGAKVTF 892

Db 833 AYNTVLAEPSENLFPASPS---MPVDGTEVTCVSGVSQKSVTCVGVGPAKSEQQVTF 888

Qy 893 LLEFEFSCSLLSQVFGKLTASDSLSLRNGTLQENTAQTSAYIQYEPHILFSSSESTLHRY 952

Db 889 TINFDFNLQNLQAAINFQAFSESQETNKA--DNSVSLTIPLLYDAELHLTRSTNIFY 946

Qy 953 EVHPYGTLP-----VGPFG-----PEFG-----TTL 972

Db 947 EISSDENAPSVIKSVEDIGPKFTFSLKVTAGSAPVSMALVTIHIPTTKENKPLLYITGI 1006

Qy 973 RTNNA---SCI--VQNLTPE---PGPPVHPPEELQHTNRLNGSNQCCQVVRCHLQGLAGT 1024

Db 1007 QTDQAGDISCTAEINPLKPLHTAPSVSFKNENFRHTKELDCRITTSKSNITCWLKDLHKA 1066

Qy 1025 EVSVGLLRLVHNEFFRRAKFKSLTVVSTFELGTTEEGSVLQLTEASRWSSESLLEVVQTRPI 1084

```
Db 1067 EYFNVTRVNRNFAASTFTQVLTAAAEIDTHNPQ-LFVIBENAVTIPLMIMKPTKA 1125
QY 1085 LISLWILGSLVGLGLLLALLVFLCWKLGF--AHKKI---PBE 1123
Db 1126 EVPTGVIIIGSIAGILLALLWAGLWKLGFQKQYKKMGQNPDE 1169

RESULT 12
Q8WY18
ID Q8WY18 PRELIMINARY; PRT; 823 AA.
AC Q8WY18; 2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MSTP018.
GN Name=MST018;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aorta;
RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
RA Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,
RA Sun Y.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
RA Gao R.L., Qiang B.O., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases;
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the Integrin alpha chain family.
DR EMBL; AF111799; AAL39001.1; .
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008305; C:integral complex; IEA.
DR GO; GO:0005155; F:protein binding; IEA.
DR GO; GO:0007160; F:cell-matrix adhesion; IEA.
DR GO; GO:0007229; F:integrin-mediated signaling pathway; IEA.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR SMART; SM00191; Int alpha; 4.
KM Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 823 AA; 92672 MW; D84E78079DCD4925 CRC64;

Query Match 23.7%; Score 1403; DB 2; Length 823;
Best Local Similarity 38.3%; Pred. No. 1.5e-91;
Matches 316; Conservative 142; Mismatches 293; Indels 74; Gaps 13;

QY 372 MSQIGFTHRLKQIGLFGMVGAYDWGGSVLWLEGGHRLFPFRMALEDEFPALQNHAYL 431
Db 1 MSQTGFSSHVVEDGVLGAGVAYDWNGAVLKETSGAKVIPLESYLKEFPPEELKNHGAYL 60
QY 432 GYSVSMLLRGGRFLPSGAPRFRHKGVIATFQLKDGAVRVAQSLOGEQISYFGSELC 491
Db 61 GYTVTSVSSRQGRVYVAGAPRFNHTKVLFTMHNRSLTTHQAMGQQIGSYFGSEIT 120
QY 492 PLDTDRDGTDLVLAAPFMFLPGQNKETGRVYVYLVQGSLLITLQGLTQPEPP-QDARFG 550
Db 121 SVDLDGQVTDVLLVAGPMYFN-EGRRGKVVYVEL-RQNFVYNGTLKOSHSQNARFG 178
QY 551 FAMGALPDLNQDGFADVAVGAPLEDGCGALYLVHGTQSGVRPHPAQRIAAASMPHALSY 610
Db 179 SSIASVRDLNQDSYNDVVVAGPLEDNHAGAIYIFHGFSGILKTPKQITASELATGLQY 238
QY 611 FGRSVDGRDLDDGDLVDVAGAGAAILLSSRPVHLTPSLEVTPOAISVVQDCRRRG 670
Db 239 FGCSTHOLDNGLDLDLAVGAGNVAIVLSRSPVQVQINASLHFPESKINIFHDKCRSG 298
QY 671 QEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAAFDGSGORLSRRLRLS 730
Db 299 RDATCLAAFLCTPIFLAPHPTQTTVGIRYNATWDERRYTPRAHLDEGGDFTRNVLVS 358
QY 731 VGNVTCQLHPLVLDTSYLPALVATVTFPDLNTPKGPVLNKGSPSTSIQKLVPFSDKCG 790
```

```
Db 359 SQQLCERINFHVLDADYVYKVPVTFVSVEYSLEDDPH-GPMLDDGMFTTLRVSPVFMNGCN 417
QY 791 PDNECVTDLVQVNMID-----RGSRK-----APFVVRGGRKKVLVST 828
Db 418 EDHCVCVDLVLDARSLPTAMEYQVRVLRKPAQDCSAYTLSPDTTTFIESTQRVAVEA 477
QY 829 TLENRKENAYNTSLSIIFSRLHLASLTPORESPIKVECAAPS--AHARLCVSGHPVFQT 886
Db 478 TLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIIECVNEBRRLQKQVCNVSYPPFRA 537
QY 887 GAKVTFLLRFPFSCSLLSQVFKLTASSDSLNRNGTLQNTAQTSAIYQYEPHLLPSSSE 946
Db 538 KAKVAFRLDFEFKSLFLHLELELAAGSDSNRSTKEDNVAAPLPHLKYEADVLFTRS 597
QY 947 STLHRYEVHPYGTLP--VGPGPEFKTTLRTN----- 975
Db 598 SSLSHVEVKLNSLSERYDGGPPFCIFRQLNGLPPIHGMMKIITPIATRSNRLKL 657
QY 976 -----NASC-IVQNLTPEPPGVHPPELQHTNRLNGSNTQCQVVVRCHLGLQAKGTEV 1026
Db 658 RDELTDEANTSCNIGWNSTERYPTPE-EDLRRAPLNHSNSDVVSINCNI-RLVPPNQEI 715
QY 1027 SVGLLRLVNEPFRRAKFKSLTVVSTFELCTEGSVLQLTASRNWSESLLEVQVTRILI 1086
Db 716 NFHLNGLMLRSLKALKYKSMKIMVNAALQRPSPFIFREEDPSRQIVFEISKQEDMQV 775
QY 1087 SLWLIGSVLGLLLALLVFLCWKLGFPAHKKIPBEEKKEKLE 1131
Db 776 PIWIVGSTLGGULLALLVLAWLKLGFFRSAR----RRREPGLD 816

RESULT 13
Q8CE84
ID Q8CE84 PRELIMINARY; PRT; 823 AA.
AC Q8CE84; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
DE library, clone:4732459H24 product:MSTP018 homolog.
GN Name=Itgall;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
```

```

RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RL prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin; DOI=10.1101/gr.152600;
RC MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuami T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirakawa T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuura T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ono M., Oheato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,
RA Saeki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akubira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AK028821; BAC26137.1; --
DR MBL; MGI:2442114; tclg1.
DR GO; GO:0006929; P:substrate-bound cell migration; IMP.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int alpha; 4.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 823 AA; 92264 MW; A330236324A0E089 CRC64;

Query Match 23.4%; Score 1388.5; DB 2; Length 823;
Best Local Similarity 38.0%; Pred. NO. 1.7e-90;
Matches 315; Conservative 138; Mismatches 296; Indels 79; Gaps 12;

QY 372 MSQIGFSTHRLKDGILFGMVCAYDWGGSVLWLEGGHRLFPFRMALEDEFFPALQNHAYL 431
DB 1 MSQIGFSSHVVEDGILLGAVGAYDWNGAVLKVETSAGKVPHPRESYLKEFFPEELKNHAYL 60
QY 432 GYSVSMLLRGGRRLFLSGAPFRHRGVIAFOLKDGAVRVAOSLQEQIGSVFGSELG 491
DB 61 GYTVSVSSVSSQGRVYVAGAPFRHNTKVILFPMHNRSLTIHQALRGEQIGSVFGSEIT 120
QY 492 PLDTDRDGTDTDLVLAAPFMFLGPQNKETGRVYVYLVGQQSLTLTQGLTQPEPP-QDARFG 550
DB 121 SVDVNDVRDVTDLVVGAPMYFS-EGRERGKVVYVNL-RQNRVYVNTGLKDSHSYQNARFG 178
QY 551 FMGALPDLNQDGFADVAVGAPLEDHGOGALYLYHGTQSGVRVHPAPQRIAAASHPALSY 610
DB 179 SCIASVQDLNDSYNDVVVVGAPLEDSDHSGRTIYIFHGFTNTLTKRPMORITASELAPGLQH 238
QY 611 FGRSVDGRLDGDGLVDVAVGAGAAITLLSSRPVHLTPSLVTPQASVVVQDCRRRG 670
DB 239 FGCSIHGQDLNEDGLVDLAVGALGNVVLWARPVQVNASLHFEPSKINIFHKDCRRNG 298
QY 671 QEAVCLTAALCFQVTSRTSPGRWDHQFYVMRFTASLDWETAGARAAFDGSGQLSPRLRLS 730
DB 299 RDATCLAAFLCFPIFLAPHPQTATVGRVYNAWTDERRYMRAHLDEGGDQFTNRAVLIS 358
QY 731 VGNVTCBQLHFLHLDTSYLRPVALTVTFFALDNTTKPGFVLNDEGSPTSIOQLVFPFSKDCG 790

```

```

Db 359 SGHEHCORINHFHLDTRADYVKPAPSVSEYSLSDPDN-GPMLDNGWPTTLRVSVFPWNGCN 417
QY 791 PDNECVTDLVQVNMDIRSRK-----APFVVRGGRKRVLPQT 828
DB 418 EDEHCVPDLVDARSDDLPTAMEYQCVLGRPAQDCSSYTLSTLSDFTTVFIESTRRRVAEA 477
QY 829 TLENKRENANTSLSIIFSRNLHLASLTPORESPIKVECAAPS--AHARLCSVGHVPQT 886
DB 478 TLENRGENAYSAVLINQSLENLOFASLIQDDSDNSIECVNEERLHKVCNVSYPFRA 537
QY 887 GAKVTFLLEPEFSCSSLLSOVFGKLTASSDSLSRNGTLQENTACTSAIOVEPHLLFSSE 946
DB 538 KAKVAFRLDFEFSKSVFLHLQHLGAGSDSHEDSTADNTALLRPHLYEADVLTFRS 597
QY 947 STLHRYEVHPYGTLPV--GPGPEFKTTLRT----- 974
DB 598 SSLSHPEVKANSSLESYDGGIGPPFNCVQNLGFFPIHGVMMKITVPIATRGGNRLIML 657
QY 975 -----NNASC-IVQNLTETPPGPPVHPELOHTNRLNGSTQCQVVRCHLGOLAKGTEV 1026
DB 658 RDPFTDQNTSCNIGWNGSTBYRSTFTB-EDLSHAPQRNHSNSDVVSIICNL-RLAPSOET 715
QY 1027 SVGLLRVHNEFFRRARFKSLTVVSTFELGTGEGSVLQLTEASRWSLSLEVVQTRPILI 1086
DB 716 SFVLVGNLWLTSLKALKYRSLKITVNAALQRPSPFIFREEDPSRQVTFEISKOEDQV 775
QY 1087 SLMLIGSVLGGLLALLVFLCKLWKGFFAHHK-----IPEEK 1125
DB 776 PIWIVGSTLGGLLALLVLMWLGFFKSARKRREPGLGPIKELK 823

RESULT 14
ITAD RAT STANDARD; PRT; 1161 AA.
AC OQYET;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE Integrin alpha-D precursor.
GN Name=Itgad;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA O'Brien M.M., VanderVliet M., Kilgannon P.D., Dietsch G.,
RA Gallatin W.M.;
RT "Cloning of rat alpha D, a novel beta 2 integrin.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and
CC VCAM1. May play a role in the atherosclerotic process such as
CC clearing lipoproteins from plaques and in phagocytosis of blood-
CC borne pathogens, particulate matter, and senescent erythrocytes
CC from the blood. (By similarity).
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D
CC associates with beta-2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

-----
CC ENBL; AF021334; AAF21241.1; -.
DR HSP; P11215; 1BHQ.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00234; WFA; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 1161 Integrin alpha-D.
FT DOMAIN 20 1100 Extracellular (Potential).
FT TRANSMEM 1101 1121 Potential.
FT DOMAIN 1122 1161 Cytoplasmic (Potential).
FT REPEAT 34 87 FG-GAP 1.
FT REPEAT 88 ? FG-GAP 2.
FT DOMAIN 152 334 WFA.
FT REPEAT 352 402 FG-GAP 3.
FT REPEAT 403 454 FG-GAP 4.
FT REPEAT 456 517 FG-GAP 5.
FT REPEAT 519 577 FG-GAP 6.
FT REPEAT 582 634 FG-GAP 7.
FT CA_BIND 467 475 Potential.
FT CA_BIND 531 539 Potential.
FT CA_BIND 594 602 Potential.
FT SITE 1126 1130 GPFKR motif.
FT DISULFID 69 76 By similarity.
FT DISULFID 108 126 By similarity.
FT DISULFID 656 711 By similarity.
FT DISULFID 769 775 By similarity.
FT DISULFID 893 860 By similarity.
FT DISULFID 993 1017 By similarity.
FT DISULFID 1022 1027 By similarity.
FT CARBOHYD 61 61 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 89 89 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 393 393 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 696 696 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 734 734 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 784 784 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 907 907 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 936 936 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;

Query Match 19.1%; Score 1135.5; DB 1; Length 1161;
Best Local Similarity 30.3%; Pred. No. 4.2e-72;
Matches 358; Conservative 203; Mismatches 483; Indels 139; Gaps 44;

Qy 20 CSPNLDEHRLPPGPPPEAFGVSVLQVGGORWMLVGAPWDPGSDRRGDVYRCVPG 79
Db 17 CHGSLNDEEPIVF-REDAASFGTVVQF--GGR-LVVGAPLEAVAVNQTGRLYDC--- 69
Qy 80 GAHNAPCAKGLHDYQLGNSHSPAVNNHGLMSLLETDDGGFMACAPLWSRACGSSVFSS 139
Db 70 ----AP-ATGMCQPIVL-RSPLEAVNMSLSLSLVTATNNAQLLACGTAQACVKNYAK 123
Qy 140 GICARVDASFPQGSFLAPTAQRCTY-MDVVIVLDGNSI--YPMSEVQTFRLRLVGLKF 196
Db 124 GSCLLGSSLQFIQAVPASPMECPQEBMDIAFLDGGSGINQRDFAQMKDFVKALMGE-F 182
Qy 197 IDPQIQVLQVYGESPVHESLGDFTKEVVRVAAKNLSRREGRETKTAQAINVACTEG 256
Db 183 ASTSTL-FSLMQYINILKTHFTTFTEFKNILDPOQLVQQL-TYTAGIKTVMEEL 240
Qy 257 FSQSHOGRPEARLLVVVVDGESH-DGSELPAALKACEAGRVTRYIGIAVLGHLRQRDP 315

```

```

-----
Db 241 FHSKNGSRSAKKILLVITDQYKRPLEYSDVI PAADKAGIIRYAIGVGDAF---QEP 296
Qy 316 SSPLRIRTIASDPDERFFPNVTDEALTDIVDALGDRIFGLEGSHAENESSGLEMSQI 375
Db 297 TA-LKELNTIGSAPPDQHVFKVGNFAALRSIQRLQEKIFAIEGTQSRSSSSQHEMSQE 355
Qy 376 GFSTHRLKDGILFGMGAYDMGGSVLWLEGHRLFPF-----RMALEDEFPALQNH 428
Db 356 GFSSALTSQGVLGAVGSPSW-----SGGARLYPNTPTPTFNNMQEN-----VDMRD 403
Qy 429 AYLGYSVSMILRGGRRLFLSGRFRHRGKVIAFQKKDGAVRAQSLQGEIGSYFGS 488
Db 404 SYLGSTAVAFWKGVHSLIL-GAPRQHTGKVIF-TQEARHWRPKSEVRGTQIGSYFGA 461
Qy 489 ELCPLDTRDGTDLVLAAPMFLGPONKETGRVYVVLV-GQOSLLTLQGTLPPEPQD- 546
Db 462 SLCSVDVDRDGTDLVLIGAPHYY--BOTRGQGVSVFPVPGVRGRWQCEATLHGEQHPW 519
Qy 547 ARFGFAMGALPDNLQDGFADVAVGAPLEQGHQALYLXHG-TQSGVRPHPAQRIAAAMP 605
Db 520 GRFGVALTVLGVNGDNLDADVAIGAPGERSRGAVIFHGASRLTMPSPSRVTSQLS 579
Qy 606 HALSYFGSRVDGRDLDDGDDLDVAVGAQGAAILSSRPVILHTPSLEVTTPQASVVQRD 665
Db 580 LRLQYFGQSLSGGQDLTQDGLVLA VGAQGHVLLRSLPLLKVELSIRPAPMEVAKAVYQ 639
Qy 666 CRRR-----GOEAVCLTAALCFQVTSRTPGRWDH-QFYMRFTASLDEWTAGARAFDG 717
Db 640 CWERTPTVLEAGBATVCLT-----VHKGSPDLLGNVQGSVRYDLALDPLRLSRAIFDE 693
Qy 718 SGQRLSPRLRLSVGNVTCBOLHFHVL-DTSDYLRPVALTPTFAL-DNTTKP---GPVLN 772
Db 694 TKNCTITGRKTLGLD-HCETVKLLIPDCVEDAVSFILILNLSLVRDSASPRNLHPVLA 752
Qy 773 EGSPTSIQKLVFPFSCDGPNECVTLVLQVNMNDRGSRKAPFVVRGGRKVLVSTTLEN 832
Db 753 VGSQDHITASLPFEKCKQELLCEGLGISFNFS-----GLQVLVVGSPDLTVTVTN 807
Qy 833 RKENAYNTSLIIFSRNLHLASLTPORES---PIKVEC-AAPSAHARL-----CSVGHVPV 884
Db 808 EGSDSYGLTVKFYYPAGLSYRRVTGTQPHQYFRLRLACEABPAAQEDLRSSSCSINHPIF 867
Qy 885 QTCAKVTFLLEPFSCSSLLSVFGKLTASSDSLERNGLQENTA-QTSAYIQV-----937
Db 868 REGAKTFMITFDVSVKAPLGDRL--LLRAKASSENKPKDPTNKTAQLELPVKYTVYTLI 925
Qy 938 -----EPHLLPSS-----ESTLHRYEYHPYGTLPVPGPPEPKTTLRTN-----975
Db 926 SRQEDSTNHNVSFSSSHGRRQEAHRYRVNLSPLKLA VRNFWPVLNGVAVMDVTLS 985
Qy 976 ----NASCIVONLTPPGPPVHPPELQHTNRLNGSNTQCVVCHLGOLAKGTEVSVGLL 1031
Db 986 SPAQGVSCVSQ--MKPPQNPDLFTQIRRSVLDCSIADCLHFRCDIPSLDIQDELDFILR 1043
Qy 1032 RLVNHEFFRAKPSLTVVSTPELGTGEGSVLQI--TEA---SRWSESLLLEVQTRPILI 1086
Db 1044 GNLSFGWVSTLQEKVLLSEAITDTSVYSQLPQEAFLRAQVETLSEYVYVEPI--1101
Qy 1087 SLMLIGSVLGGILLALLVLCMLKGLFFAHKKIPIBEEREK 1129
Db 1102 --FLVAGSSVGGILLALLITVLYVLYKLGFP-----KQYK 1133

RESULT 15
ITAD HUMAN
ID ITAD HUMAN STANDARD; PRT; 1162 AA.
AC Q13349; Q15575; Q15576;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-D precursor (leukointegrin alpha D) (CD11d) (ADB2).
GN Name=ITGAD;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 143.162 Seconds
(without alignments)
3058.161 Million cell updates/sec

Title: US-09-647-544-4
Perfect score: 5932
Sequence: 1 MELPFTVHLPLVFLTGLC.....GFFAHKKIPREKREKLEQ 1132

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_16Dec04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5932	100.0	1132	3	AAY32243 Human int
2	5904.5	99.5	1167	3	AAY32242 Human int
3	5890.5	99.3	1167	8	ADQ19290 Human sof
4	5838.5	98.4	1167	4	AAB64584 Human sec
5	5838.5	98.4	1167	6	ABP99490 Human sec
6	5838.5	98.4	1167	6	ABR00964 Human gen
7	5838.5	98.4	1167	6	ADA44026 Human sec
8	5829.5	98.3	1152	4	AAB64657 Human sec
9	5763.5	97.2	1152	4	AAB64658 Human sec
10	4878.5	82.2	1049	5	ADR41424 Human CD-
11	2355	39.7	1188	4	AAU14231 Human nov
12	2355	39.7	1188	4	AAU14231 Human nov
13	2355	39.7	1188	5	AAU10551 Human A25
14	2354	39.7	1188	4	AAU14467 Human nov
15	2354	39.7	1188	7	ADE09956 Novel pro
16	2344.5	39.5	1189	3	AAU12582 ITGAl1 pr
17	2344.5	39.5	1189	4	ABG12949 Novel hum
18	2344.5	39.5	1189	6	ABR58364 Human NOV
19	2344.5	39.5	1189	6	ADA27054 Human nov
20	2344.5	39.5	1189	7	ADE63570 Human Pro
21	2344.5	39.5	1189	8	ADE86584 Novel hum
22	2344.5	39.5	1189	8	ADQ19968 Human sof
23	2337.5	39.4	1188	4	AAU50087 Murine A2
24	2337.5	39.4	1188	5	AAU10552
25	2319	39.1	1188	4	AAU30929 Amino aci

26	2269.5	38.3	1189	8	ADH80785	Adh80785 Human pol
27	2194	37.0	1034	3	AAB25590	Aab25590 Protein e
28	2194	37.0	1034	6	ADA27062	Ada27062 Human nov
29	2194	37.0	1034	8	ADE86592	Ade86592 Novel hum
30	2194	37.0	1058	5	ADR41496	Adr41496 Human CD-
31	2140	36.1	1120	6	ABR58365	AbR58365 Human NOV
32	1979	33.4	987	8	ADP29492	Adp29492 Human sec
33	1887.5	31.8	1179	5	ABB90759	Abb90759 Human Tum
34	1887.5	31.8	1179	5	ABP64915	Abp64915 Human pro
35	1887.5	31.8	1179	6	ABU54466	Abu54466 Human tum
36	1884	31.8	1180	5	ABR90788	AbR90788 Rat Tumou
37	1884	31.8	1180	6	ABU54495	Abu54495 Mouse tum
38	1884	31.8	1180	7	ADE63568	Ade63568 Rat Prote
39	1881.5	31.7	1151	8	ADE86652	Ade86652 Human int
40	1874.5	31.6	1177	8	ADQ39498	Adq39498 Human myo
41	1825.5	30.8	1183	2	AAU07728	Aau07728 Armenian
42	1733.5	29.2	1181	6	ABU03548	Abu03548 Angiogene
43	1733.5	29.2	1181	6	ABR59703	AbR59703 Human VLA
44	1733.5	29.2	1181	6	AAG79775	Aag79775 Alpha2 in
45	1733.5	29.2	1181	6	ABU03616	Abu03616 Human exp

ALIGNMENTS

RESULT 1
AAY32243
ID AAY32243 standard; protein; 1132 AA.
XX
AC AAY32243;
XX
AC AAY32243;
DT 15-FEB-2000 (first entry)
XX
DE Human integrin subunit alpha-10 splice variant.
XX
KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;
KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker;
KW splice variant.
XX
OS Homo sapiens.
XX
FH Location/Qualifiers
FT Peptide
FT i. .22
FT /note= "signal peptide"
FT Protein
FT 23. .1132
FT /note= "mature protein"
XX
XX
WO9951639-A1.
XX
PD 14-OCT-1999.
XX
XX
PF 31-MAR-1999; 99WO-SE000544.
XX
XX
PR 02-APR-1998; 98SE-00001164.
PR 28-JAN-1999; 99SE-00000319.
XX
XX
PA (ACTI-) ACTIVE BIOTECH AB.
XX
XX
PI Lundgren-Akerlund E;
XX
XX
DR WPI; 2000-052639/04.
DR N-PSDB; AAZ34720.
XX
XX
PT New isolated integrin subunit alpha-10, used as a marker or target
PT molecule for cells during development, regeneration and pathological
PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
PT inflammation.
PS
PS Claim 1; Page 43-48; 90pp; English.
XX
XX This sequence represents a splice variant of novel human chondrocyte
CC integrin subunit alpha-10 (ISa10). It is identical to ISa10 (see

Db 1141 VFCLWKLGFPAHKIPBEKREKLEQ 1167

RESULT 3

ADQ19290

ID ADQ19290 standard; protein; 1167 AA.

XX AC ADQ19290;

XX DT 26-AUG-2004 (first entry)

XX DE Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.

XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

XX OS Homo sapiens.

XX PN WO2004048938-A2.

XX PD 10-JUN-2004.

XX PF 26-NOV-2003; 2003WO-US038193.

XX PR 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX PA Aziz N, Ginsburg WM, Zlotnik A;

XX PI WPI; 2004-441208/41.

XX PT Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.

XX PS Example 2; SEQ ID NO 2109; 210pp; English.

XX CC The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytosstatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated protein of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

XX SQ Sequence 1167 AA;

Query Match 99.3%; Score 5890.5; DB 8; Length 1167;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1129; Conservative 2; Mismatches 1; Indels 35; Gaps 1;

Qy 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFGPPPEABRGYSVLQHVGGQRMVLGA 60

Db 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFGPPPEABRGYSVLQHVGGQRMVLGA 60

Qy 61 PWDGPGDRRDGVYRCPVGGAHNAPCAKGLGDYQLGNSSHPAVNMHLGMSLLETGDDGG 120

Db 61 PWDGPGDRRDGVYRCPVGGAHNAPCAKGLGDYQLGNSSHPAVNMHLGMSLLETGDDGG 120

Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCTYMDVIVLDGNSIYP 180

Db 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCTYMDVIVLDGNSIYP 180

Qy 181 WSEVQTFLLRLVGLKFLIDPEIQVLQVYGSPVHWSLGDGFRKKEEVRAAKNLSRREG 240

Db 181 WSEVQTFLLRLVGLKFLIDPEIQVLQVYGSPVHWSLGDGFRKKEEVRAAKNLSRREG 240

Qy 241 RETKTAQAIWVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKACAGRVTRY 300

Db 241 RETKTAQAIWVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKACAGRVTRY 300

Qy 301 GIAVLGHYLRQRDPSSFLREINTIASDDPDERFFNVNTEAALTIDVLDGDRIFGLEGS 360

Db 301 GIAVLGHYLRQRDPSSFLREINTIASDDPDERFFNVNTEAALTIDVLDGDRIFGLEGS 360

Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGGHRLPFPPMALEDEF 420

Db 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGGHRLPFPPMALEDEF 420

Qy 421 PPALQNHAAVLYGYSVSSMLRGRRRLFLSGAPFRHRGKVIAFQLKKDGAVRVAQSLQGE 480

Db 421 PPALQNHAAVLYGYSVSSMLRGRRRLFLSGAPFRHRGKVIAFQLKKDGAVRVAQSLQGE 480

Qy 481 QIGSYFGSELCPDTRDGTVDVLLVAAPMFLGPQNKETGRVYVYLVGQSLTLTQGTLLQ 540

Db 481 QIGSYFGSELCPDTRDGTVDVLLVAAPMFLGPQNKETGRVYVYLVGQSLTLTQGTLLQ 540

Qy 541 PEPQDARFGFAMGALPDLNODGFADVAVGAPLEDGHQALYLYHGTQSGVRPHPAQRIA 600

Db 541 PEPQDARFGFAMGALPDLNODGFADVAVGAPLEDGHQALYLYHGTQSGVRPHPAQRIA 600

Qy 601 AASMPHALSYFGRSVDGRDLDDGDDLDVAVAGAGAAIILSSRPVHLTPSLEVTPOAIS 660

Db 601 AASMPHALSYFGRSVDGRDLDDGDDLDVAVAGAGAAIILSSRPVHLTPSLEVTPOAIS 660

Qy 661 VWORDCRRRQBAVCLTAALCFQVTSRTPGRWDPHQYMFRTASLDEWTAGARAAFDGSGQ 720

Db 661 VWORDCRRRQBAVCLTAALCFQVTSRTPGRWDPHQYMFRTASLDEWTAGARAAFDGSGQ 720

Qy 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALTVPFALDNTTKPGPVNLNGSPSTIQ 780

Db 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALTVPFALDNTTKPGPVNLNGSPSTIQ 780

Qy 781 KLVFPSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGGRKVLVSTTLLENKENAYNT 840

Db 781 KLVFPSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGGRKVLVSTTLLENKENAYNT 840

Qy 841 SLSLIISRNHLASLTQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLEPEFSC 900

Db 841 SLSLIISRNHLASLTQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLEPEFSC 900

Qy 901 SSLLSQVFGKLTASSDSLEBNGTLQNTAQTSAYIOVEPHLLFSSSTLHRYEHPVGTLL 960

Db 901 SSLLSQVFGKLTASSDSLEBNGTLQNTAQTSAYIOVEPHLLFSSSTLHRYEHPVGTLL 960

Qy 961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985

Db 961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985

Qy 986 EPPGPPVHPPELOHTNRLNGSTQCQVVRCHLQOLAKGTSEVSGLLRLVNEFFRRAPFK 1045

Db 1021 EPPGPPVHPPELOHTNRLNGSTQCQVVRCHLQOLAKGTSEVSGLLRLVNEFFRRAPFK 1080

Qy 1046 SLTVVSTFELGTSEGSVLQLTEASRWSESLEVVQTRPILISWILIGSVLGLLALL 1105

Db 1081 SLTVVSTFELGTSEGSVLQLTEASRWSESLEVVQTRPILISWILIGSVLGLLALL 1140

Qy 1106 VFCLWKLGFPAHKIPBEKREKLEQ 1132

Db 1141 VFCLWKLGFPAHKIPBEKREKLEQ 1167

RESULT 4

AAB64584

ID AAB64584 standard; protein; 1167 AA.

XX AC AAB64584;

XX DT 22-MAR-2001 (first entry)

Human; secreted protein; nontropic; neuroprotective; cytostatic; virucide; dermatological; immunosuppressive; anti-inflammatory; anti-viral; antibacterial; antiparkinsonian; antitickling; antianemic; antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective; antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine; cardiovascular disorder; neurological disease; nephrotropic; gene therapy.

OS Homo sapiens.

XX W0200277186-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002WO-US009188.

XX 27-MAR-2001; 2001US-0278650P.

PR 12-SEP-2001; 2001US-00950082.

PR 12-SEP-2001; 2001US-00950083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-040583/03.

DR N-PSDB; ABZ66911.

XX New human secreted proteins encoded by genes contained in cDNA clones (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS, multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or West Nile fever.

XX Claim 1; Page 1395-1398; 2423pp; English.

XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the encoded secreted proteins (ABP99470-ABP99872) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections

XX Sequence 1167 AA;

Query Match 98.4%; Score 5838.5; DB 6; Length 1167;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 1120; Conservative 3; Mismatches 9; Indels 35; Gaps 1;

QY 1 MELPFVTHLFLVFLVTLGLCSFNLDEHRLPFGPEAEFGYSLVGVGGQRMVLGA 60
DB :|||||
1 VELPFVTHLFLVFLVTLGLCSFNLDEHRLPFGPEAEFGYSLVGVGGQRMVLGA 60
QY 61 PWGPGSDRRGDVYRCPVGAHAPCAKGLGDYQLGNSHPVNMHLGMSLLETDDGG 120
DB 61 PWGPGSDRRGDVYRCPVGAHAPCAKGLGDYQLGNSHPVNMHLGMSLLETDDGG 120
QY 121 FMACAPLWSRACGSVSSGICARVDASFPQGS LAPTAQRCPYMDVIVLDGNSIYP 180
DB 121 FMACAPLWSRACGSVSSGICARVDASFPQGS LAPTAQRCPYMDVIVLDGNSIYP 180
QY 181 WSEVQTFRLRLVGLFDIDPQIOGLVQYGESPVHWSLGDFTKEEVRAAKNLSREG 240
DB 181 WSEVQTFRLRLVGLFDIDPQIOGLVQYGESPVHWSLGDFTKEEVRAAKNLSREG 240

QY 241 RETKTAQAIMVACTEGFSQSHGQRPAAARLLVVVTGDESHGDBELPAALKACAGRVTRY 300
DB |||||
241 RETKTAQAIMVACTEGFSQSHGQRPAAARLLVVVTGDESHGDBELPAALKACAGRVTRY 300
QY 301 GIAVLGHYLRQRDPSPFLREIRTIASDDPERFFNVNTEAALTDIVDALGDRIFGLEGS 360
DB |||||
301 GIAVLGHYLRQRDPSPFLREIRTIASDDPERFFNVNTEAALTDIVDALGDRIFGLEGS 360
QY 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGHRLPPPMALDEDEF 420
DB |||||
361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGHRLPPPMALDEDEF 420
QY 421 PPALQNHAAAYLYGYSVSMLLRGRRFLSGAPFRHRGKVIAFQLKKGAVRVAQSLQGE 480
DB |||||
421 PPALQNHAAAYLYGYSVSMLLRGRRFLSGAPFRHRGKVIAFQLKKGAVRVAQSLQGE 480
QY 481 QIGSYFGSELCLDTRDGTDDVLLVAAPMFLGPONKGTGRVVYLVGQSSLLTLOQTLO 540
DB |||||
481 QIGSYFGSELCLDTRDGTDDVLLVAAPMFLGPONKGTGRVVYLVGQSSLLTLOQTLO 540
QY 541 PEPPQDARFGFAMGALPDLNODGFADVAVGAPLEDGHQGALYLYHGTQSGVRPHPAORIA 600
DB |||||
541 PEPPQDARFGFAMGALPDLNODGFADVAVGAPLEDGHQGALYLYHGTQSGVRPHPAORIA 600
QY 601 AASMPHALSYFGRSVDGRDLDDGDDLVDAVAGAGAAIILSSRPVHLTPSLEVTPOAIS 660
DB |||||
601 AASMPHALSYFGRSVDGRDLDDGDDLVDAVAGAGAAIILSSRPVHLTPSLEVTPOAIS 660
QY 661 VWORDCRRQGEAVCLTAALCFQVTSRTQGRWDHOFYMRFTASLDEWTAGARAAFDGSGQ 720
DB |||||
661 VWORDCRRQGEAVCLTAALCFQVTSRTQGRWDHOFYMRFTASLDEWTAGARAAFDGSGQ 720
QY 721 RLSPLRLRLSVGNVTCQLHFFHVLDTSDYLRPVALTVPALDNTTKPGPVNLGSGSPSIQ 780
DB |||||
721 RLSPLRLRLSVGNVTCQLHFFHVLDTSDYLRPVALTVPALDNTTKPGPVNLGSGSPSIQ 780
QY 781 KLVFPKDCGPDNECVTDLVQVNMDIRSKAPFVVRGGRKRVLVSTTLLENKENAYNT 840
DB |||||
781 KLVFPKDCGPDNECVTDLVQVNMDIRSKAPFVVRGGRKRVLVSTTLLENKENAYNT 840
QY 841 SLSIISRNHLASLPQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLLEPFSC 900
DB |||||
841 SLSIISRNHLASLPQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLLEPFSC 900
QY 901 SSLLSVOFGKLTASSDLSRNGTLQNTAQTSAIOVEPHLLFSSESTLHRYEVPYGTLL 960
DB |||||
901 SSLLSVOFGKLTASSDLSRNGTLQNTAQTSAIOVEPHLLFSSESTLHRYEVPYGTLL 960
QY 961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985
DB |||||
961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985
QY 986 EPGGPPVHPELOHTNRLNGSNTCCQVVRCHLQOLAKGTEVSGLLRLVNEFFRRKAPK 1045
DB |||||
1021 EPGGPPVHPELOHTNRLNGSNTCCQVVRCHLQOLAKGTEVSGLLRLVNEFFRRKAPK 1080
QY 1046 SLTVVSTFEIQTGEGSVLQTEASRWSESLLEVVTQTRPILISIMILIGSVLGLLALL 1105
DB |||||
1081 SLTVVSTFEIQTGEGSVLQTEASRWSESLLEVVTQTRPILISIMILIGSVLGLLALL 1140
QY 1106 VFCLMKLGFPAHKKIPPEEKREKLEQ 1132
DB |||||
1141 VFCLMKLGFPAHKKIPPEEKREKLEQ 1167
RESULT 6
ABR00964
ID ABR00964 standard; protein; 1167 AA.
XX ABR00964;
XX AC
XX DT 12-MAY-2003 (first entry)
XX

Human gene 18-encoded secreted protein HAOAG15, SEQ ID NO:445.

Human; secreted protein; cancer; tumour; hyperproliferative disorder; autoimmune disorder; inflammation; angioinvasive diseases; AIDS; acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing; drug screening; chromosome identification; chromosome mapping; cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV; antitumour; vulnery; chromosome 1q21.

OS Homo sapiens.

XX PN WO200277013-A2.

XX
PD
03-OCT-2002.

XX PF 26-MAR-2002: 2002WO-US009370.

XX
PR 27-MAR-2001: 2001US-0278650P

PR 12-SEP-2001; 2001US-009500B2.
PR 12-SEP-2001; 2001US-009500B3.

XX PA (HITMA-) HITMAN GENOME SCT TNC

XX	Page CM.
DT	Page CM.

XX
MDT. 2003-040579/03

DR N-PSDB; ABZ73298.

PT New human secreted

PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.

PS Claim 13; Page 1387-1390; 2474pp; English.

AB273281-AB273697 represent cDNAs corresponding to 391 human secreted protein genes, and AB00947-AB03163 represent the proteins they encode. The AB273698-AB274687 represent human secreted protein genomic fragments. The invention also encompasses antibodies specific for the secreted proteins, the use of the secreted proteins in drug screening and recombinant vectors and host cells comprising a nucleic acid of the invention. The secreted proteins are thought to be involved in biological activities associated with cellular signalling, cellular differentiation, cell migration, prohormone activation and neurotransmitter activity. The secreted proteins, nucleic acids encoding them, antibodies or antibody fragments specific for the secreted proteins, and modulators of protein activity are useful for diagnosing or treating cancers or other hyperproliferative disorders. Additionally, the secreted proteins and their nucleic acids may also be used in the treatment of autoimmune disorders, inflammatory disorders, diseases involving angiogenesis, AIDS (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote wound healing. Nucleic acids of the invention may be used for chromosome identification, chromosome mapping, in gene therapy, for identifying individuals from minute biological samples, as hybridisation probes, and as molecular weight markers. The present sequence represents a human secreted protein of the invention

Sequence 1167 AA;

Query Match	98.4%	Score 5338.5;	DB 6;	Length 1167;
Best Local Similarity	96.0%;	Pred. No. 0;		
Matches 1120; Conservative	3;	Mismatches	9;	Indels 35; Gaps 1;

QY 1 MELPFVTHLFLPLVFLTGLCSPFNLDHHHPRLFPGPPEAEFGYSVLQHVGGQRWMLVGA 60

```

db
1 VELPFVTHLEPLVFTGLCSFFNLDEHHPRLFPPPPPEAEFGYSVLOHVGGGORWMLVGA 60
:
```

QV 61 PWDGPGDRRGDVYRCPVGGAHNAPCAKGHLGDYOLGNSSHPAVNMHLGMSLLETGDGG 120

db 61 PWDGSGDRRGDPYRCVCGAHNAPCAKCHLGDYOLGNSSHPAVNMHLGMSLLETDPDGG 120

121 FMACAPLWSRACGSSVFSSGICARVDASEFOPOGSLAPTAQRCPITYMDVIVL.DGNSNIYP 180

121 FMACAPLWSRACGSSVFSSGICARVDSFOGSLAPTAORCPTYPDVIVLDGNSNIYP 180

Qy	181	WSEVQTFRLRLVCKLFIDPEIQIUGLVQVGESPVHWSLGDFTKEEVRPAAKNLSRREG	240
Db	181	WSEVQTFRLRLVCKLFIDPEIQIUGLVQVGESPVHWSLGDFTKEEVRPAAKNLSRREG	240
Qy	241	RETAKTAQIMVACTEGFSQSHGCRPEAARLLVVVTDGESHDGBELPAALKACBAGRVTY	300
Db	241	RETAKTAQIMVACTEGFSQSHGCRPEAARLLVVVTDGESHDGBELPAALKACBAGRVTY	300
Qy	301	GI AVLGHYLRQRDPSPSFUREIRTIASDPDERFFNVNTDEAALTDIIDVALGDRIFGLEGS	360
Db	301	GI AVLGHYLRQRDPSPSFUREIRTIASDPDERFFNVNTDEAALTDIIDVALGDRIFGLEGS	360
Qy	361	HANESSFGLEMSQIGFSTHRLKXGILFGMWGAYDMGGSVLWLEGGHRLFPFRMALEDEF	420
Db	361	HANESSFGLEMSQIGFSTHRLKXGILFGMWGAYDMGGSVLWLEGGHRLFPFRMALEDEF	420
Qy	421	PPALQNHAAIYIGYSVSMLLRGRRRLFLSCAPFRHRGKVIATQFKDKDGAVRVAQSLQGE	480
Db	421	PPALQNHAXIYIGYSXSMULLRGXRLLXSGAXFRHRGKVIATQFKDKDGAVRVAQSLQGE	480
Qy	481	QIGSYFGSELCPIDTRDGTDDTVLLVAAPMFLGPONKETGRVVYLVGQOSLLTLOQTLO	540
Db	481	QIGSYFGSELCPIDTRDGTDDTVLLVAAPMFLGPONKETGRVVYLVGQOSLLTLOQTLO	540
Qy	541	PEPPQDARFCFANGALPDLNODGCFADVAVGAPLEDHOGALYLYVHGTYOSGVRPHPAQRIA	600
Db	541	PEPPQDARFCFANGALPDLNODGCFADVAVGAPLEDHOGALYLYVHGTYOSGVRPHPAQRIA	600
Qy	601	AASMPHALSYFGRSDVGRDLDDGDDLVDAVGAQGAAILLSRPIVHLTPSLEVTPQAIS	660
Db	601	AASMPHALSYFGRSDVGRDLDDGDDLVDAVGAQGAAILLSRPIVHLTPSLEVTPQAIS	660
Qy	661	VVQDCCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAADFQSGQ	720
Db	661	VVQDCCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAADFQSGQ	720
Qy	721	RLSPRLRLSVGNNVTCBQLHFHVLDTSDYLRPVALTVTFALDNTTKPGPVLNEGSPTSIQ	780
Db	721	RLSPRLRLSVGNNVTCBQLHFHVLDTSDYLRPVALTVTFALDNTTKPGPVLNEGSPTSIQ	780
Qy	781	KLVPFCKDCPDNECVTDVLQVNMDDIRGSRKAPFVVRRGRRKVLVSTTTLENRKENAYNT	840
Db	781	KLVPFCKDCPDNECVTDVLQVNMDDIRGSRKAPFVVRRGRRKVLVSTTTLEXRKENAYNT	840
Qy	841	SLSLIISRNHLASLTPQRESPIKVECAAPSAAHARLCSVGHPPVQTCAKYTFELFEFFSC	900
Db	841	SLSLIISRNHLASLTPQRESPIKVECAAPSAAHARLCSVGHPPVQTCAKYTFELFEFFSC	900
Qy	901	SSLLSQVFGKLTASSDSLENGTLQENTAQTSAYIOYEPHLLFSSESTLHRYEYHPYCTL	960
Db	901	SSLLSQVFKVLTAASSDSLENGTLQDNTAQTSAYIOYEPHLLFSSESTLHRYEYHPYCTL	960
Qy	961	PVGGPBEFKTTLR-----TNNASCIQONLT	985
Db	961	PVGGPBEFKTTLRVQLNLCVWVWSGLIISALLPAVAHGNYFLSLSQVITNNASCIQONLT	1020
Qy	986	EPGPPVHPELOHTNPLNGSNTQCQVVRCHLGOLAKTEVSVGLRLVHNEFFRRRAKFK	1045
Db	1021	EPGPPVHPELOHTNPLNGSNTQCQVVRCHLGOLAKTEVSVGLLRLVHNEFFRRRAKFK	1080
Qy	1046	SLTVVSTFELGTBEGSVLQLTEASRWSESLEVVQTRPILISLWILGSLVGLGLLALL	1105
Db	1081	SLTVVSTFELGTBEGSVLQLTEASRWSESLEVVQTRPILISLWILGSLVGLGLLALL	1140
Qy	1106	VFLCWKLGFFFAHKKIPEEBKREKLEQ	1132
Db	1141	VFLCWKLGFFFAHKKIPEEBKREKLEQ	1167

RESULT 7

ADA44026

ID ADA44026 standard; protein; 1167 AA.

XX ADA44026;
XX AC
XX 20-NOV-2003 (first entry)
XX DE Human secreted protein SEQ ID 214.
XX KW Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;
XX KW Neuroprotective; Cerebroprotective; Antianemic.
XX OS Homo sapiens.
XX PN WO2003000865-A2.
XX PD 03-JAN-2003.
XX PF 26-MAR-2002; 2002WO-US009105.
XX PR 27-MAR-2001; 2001US-0278650P.
XX PR 12-SEP-2001; 2001US-00950082.
XX PR 12-SEP-2001; 2001US-00950083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR N-PSDB; ADA43832.
XX FT A human secreted protein and nucleic acids useful for preparing a
XX PT diagnostic or pharmaceutical composition for diagnosing or treating
XX PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,
XX PT retinopathy, neuropathy.
XX PS Claim 1; SEQ ID NO 214; 701pp; English.
XX CC The invention relates to novel genes and their fragments which are useful
XX CC for preventing, treating or ameliorating medical conditions e.g. by
XX CC protein or gene therapy. The genes are isolated from a range of human
XX CC tissues disclosed in the specification. The nucleic acids and proteins
XX CC are useful in the diagnosis, treatment and prevention of conditions
XX CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,
XX CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,
XX CC infection, cataract, renal disorders, or endocrine disorders. The present
XX CC sequence was used to illustrate the invention.
XX SQ Sequence 1167 AA;

Query Match 98.4%; Score 5838.5; DB 6; Length 1167;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 1120; Conservative 3; Mismatches 9; Indels 35; Gaps 1;

Qy 1 MELPFVTHLFLPLVFLTGLCSPPFNLDHHPRLFPGPPEAEFGYSVLQHVGGQRWMLVGA 60
Dy :
Dy 1 VELPFVTHLFLPLVFLTGLCSPPFNLDHHPRLFPGPPEAEFGYSVLQHVGGQRWMLVGA 60

Qy 61 PWDGPGDGRGDVRCVPVGGAHNAPCAKHGLDYLQGNSSHPANMHLGMSLLETDDGG 120
Dy :
Dy 61 PWDGPGDGRGDVRCVPVGGAHNAPCAKHGLDYLQGNSSHPANMHLGMSLLETDDGG 120

Qy 121 FMACAPLMSRACGSSVSSGICARVDASFPQGS LAPTAQRCPTMYDVIVLDGNSIYP 180
Dy :
Dy 121 FMACAPLMSRACGSSVSSGICARVDASFPQGS LAPTAQRCPTMYDVIVLDGNSIYP 180

Qy 181 WSEVQTFLLRLVGLKFLTDPEIQVGLVQYGSPPHWSLGDPRFKKEEVVRAAKNLSRREG 240
Dy :
Dy 181 WSEVQTFLLRLVGLKFLTDPEIQVGLVQYGSPPHWSLGDPRFKKEEVVRAAKNLSRREG 240

Qy 241 RETTQAQIMVACTEGFSQSHGGRPEARLAVVTTDGHSHDGEELPAALKAACEAGRVTRY 300
Dy :
Dy 241 RETTQAQIMVACTEGFSQSHGGRPEARLAVVTTDGHSHDGEELPAALKAACEAGRVTRY 300

Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360

Db 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVAGYDMGGSVLMEGGHRLPPPMALDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVAGYDMGGSVLMEGGHRLPPPMALDEF 420
Qy 421 PPALQNHAAVLYGYSVSMILRGRRRLPLSGAPFRHRGKVIAPQLKDGAVRVAQSLOGE 480
Db 421 PPALQNHAAVLYGYSVSMILRGRRRLPLSGAPFRHRGKVIAPQLKDGAVRVAQSLOGE 480
Qy 481 QIGSYFSGELCPLDTRDGTDLVLAAPMFLGPQNKETGRVYVYLVGQSLTLTQGLTQ 540
Db 481 QIGSYFSGELCPLDTRDGTDLVLAAPMFLGPQNKETGRVYVYLVGQSLTLTQGLTQ 540
Qy 541 PEPQDARFGFANGALPDNLNODGFADVAVGAPLEDHOGALYLYHGTQSGVRPAPQRIA 600
Db 541 PEPQDARFGFANGALPDNLNODGFADVAVGAPLEDHOGALYLYHGTQSGVRPAPQRIA 600
Qy 601 AASMPHALSYFGRSVDGRDLDDGDDLDVAVAGAAAILSSRPVHLTPSLEVTQAIS 660
Db 601 AASMPHALSYFGRSVDGRDLDDGDDLDVAVAGAAAILSSRPVHLTPSLEVTQAIS 660
Qy 661 VWORDCRRRQGEAVCLTAALCFQVTSRTPGRWDPHFYMRPTASLDEWTAGARAAFDGSGQ 720
Db 661 VWORDCRRRQGEAVCLTAALCFQVTSRTPGRWDPHFYMRPTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSVGNVTCQLHGFHVLDTSDYLRPVALTVPFALDNTTKPGPVLNBSGSPSIQ 780
Db 721 RLSPRRLRLSVGNVTCQLHGFHVLDTSDYLRPVALTVPFALDNTTKPGPVLNBSGSPSIQ 780
Qy 781 KLVPFSDKDCPDNECVTDLVQNMDIRGSRKAPFVVRGRRKVLVSTTLENKENAYNT 840
Db 781 KLVPFSDKDCPDNECVTDLVQNMDIRGSRKAPFVVRGRRKVLVSTTLENKENAYNT 840
Qy 841 SLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLEPEFSC 900
Db 841 SLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLEPEFSC 900
Qy 901 SSLLSQVFGKLTASSSLERNGLQNTAQTSAIYQEPHLLFSSSTLHRYEVHPYGTIL 960
Db 901 SSLLSQVFGKLTASSSLERNGLQNTAQTSAIYQEPHLLFSSSTLHRYEVHPYGTIL 960
Qy 961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985
Db 961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985
Qy 986 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLQOLAKGTSEVSVGLLRLVHNEFFRAKFK 1045
Db 1021 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLQOLAKGTSEVSVGLLRLVHNEFFRAKFK 1080
Qy 1046 SLTVSTFELGTGEGSVLQLTASRSSESLLEVVQTRPILISILMILIGSVLGGILLALL 1105
Db 1081 SLTVSTFELGTGEGSVLQLTASRSSESLLEVVQTRPILISILMILIGSVLGGILLALL 1140
Qy 1106 VFCLMKLGFFPAHKKIPEEKREKLEQ 1132
Db 1141 VFCLMKLGFFPAHKKIPEEKREKLEQ 1167

RESULT 8
AAB64657
ID AAB64657 standard; protein; 1152 AA.
XX AAB64657;
XX AC AAB64657;
XX DT 22-MAR-2001 (first entry)
XX DE Human secreted protein BLAST search protein SEQ ID NO: 167.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;

KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX Homo sapiens.
 XX WO200077197-A1.
 XX 21-DEC-2000.
 XX 01-JUN-2000; 2000MO-US014934.
 XX 11-JUN-1999; 99US-0138599P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX Rosen CA, Ruben SM, Komatsoulis GA;
 XX WPI; 2001-032312/04.
 XX Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 XX Disclosure; Page 543-546; 558pp; English.
 CC The invention relates to the isolation of genes AAF32757-F32803 encoding
 CC the human secreted proteins AAB64549-B64594. The sequence is a search
 CC result from a BLASTX homology search. The genes and proteins are useful
 CC for preventing, ameliorating or treating medical conditions, e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment and
 CC prevention of: (a) cancer, e.g. breast and ovarian cancer, and other
 CC cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal
 CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
 CC disease, allergies, autoimmune haemolytic anaemia, autoimmune
 CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
 CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders
 CC such as myocardial ischaemia; (d) wound healing; (e) neurological
 CC diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases
 CC such as viral, bacterial, fungal and parasitic infections
 XX Sequence 1152-AA;

Query Match 98.3%; Score 5829.5; DB 4; Length 1152;
 Best Local Similarity 97.0%; Pred. No. 0;
 Matches 1117; Conservative 0; Mismatches 0; Indels 35; Gaps 1;

QY 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPQPPPEAERFGYSVLQHVGGQRWMLVGA 60
 DB 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPQPPPEAERFGYSVLQHVGGQRWMLVGA 60
 QY 61 PWDGPGDRRGDVVRCVPGGAHAPCAKHGLGDYQLGNSSHPAVNMHLGMSLLTDDGG 120
 DB 61 PWDGPGDRRGDVVRCVPGGAHAPCAKHGLGDYQLGNSSHPAVNMHLGMSLLTDDGG 120
 QY 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTMYDVIVLDGNSNIYP 180
 DB 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTMYDVIVLDGNSNIYP 180
 QY 181 WSEVQTFELRLVGLKFLDIPDQIOVLQVQVGSPPVHWSLGDPRFKEEVVRAAKNLSRREG 240
 DB 181 WSEVQTFELRLVGLKFLDIPDQIOVLQVQVGSPPVHWSLGDPRFKEEVVRAAKNLSRREG 240
 QY 241 RETTQAQIMVACTEGFSQSQSGHGPPEARLAVVTDGESHDPGEELPAALKACEAGRTRY 300
 DB 241 RETTQAQIMVACTEGFSQSQSGHGPPEARLAVVTDGESHDPGEELPAALKACEAGRTRY 300
 QY 301 GIAVLGHLRQRDPSSFLRIRTIASDPDRFFFNVTDEAALTDIVDALGDRIFGLGGS 360
 DB 301 GIAVLGHLRQRDPSSFLRIRTIASDPDRFFFNVTDEAALTDIVDALGDRIFGLGGS 360
 QY 361 HAENESSFLEMSQIGFSTHRLKDGILFGMVGYDWCGLVWLEGHRLFPFRMALEDEF 420

DB 361 HAENESSFLEMSQIGFSTHRLKDGILFGMVGYDWCGLVWLEGHRLFPFRMALEDEF 420
 QY 421 PPALQNHAAAYLGYSVSSMLLRGGRRLFLSGAPRFRHGRKVIAFOLKKGAVRVAQSLQGE 480
 DB 421 PPALQNHAAAYLGYSVSSMLLRGGRRLFLSGAPRFRHGRKVIAFOLKKGAVRVAQSLQGE 480
 QY 481 QIGSYFSGSELCPDTRDGGTDLVLAAPMFLGPONKGTGRVYVYLQGOSSLLTLOGLTLO 540
 DB 481 QIGSYFSGSELCPDTRDGGTDLVLAAPMFLGPONKGTGRVYVYLQGOSSLLTLOGLTLO 540
 QY 541 PEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDGCHQGALYLYHGTQSGVRPAPQRIA 600
 DB 541 PEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDGCHQGALYLYHGTQSGVRPAPQRIA 600
 QY 601 AASMPHALSYFGRSVDCGRDLDDGDDLDVAVAGAAIILSSRPVHLTPSLEVTPOAIS 660
 DB 601 AASMPHALSYFGRSVDCGRDLDDGDDLDVAVAGAAIILSSRPVHLTPSLEVTPOAIS 660
 QY 661 VVQDCCRRRQGEAVCLTAALCFQVTSRTGCRWDHOFYMRFTASLDEWTAGARAAFDGSGQ 720
 DB 661 VVQDCCRRRQGEAVCLTAALCFQVTSRTGCRWDHOFYMRFTASLDEWTAGARAAFDGSGQ 720
 QY 721 RLSPPRLRLSVGNVTCBOLHFHVLDTSDYLRPVALTVTFALDNTTKPGPVLNEGSPTSIO 780
 DB 721 RLSPPRLRLSVGNVTCBOLHFHVLDTSDYLRPVALTVTFALDNTTKPGPVLNEGSPTSIO 780
 QY 781 KLVPFSSKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKRVLSTTLENRKENAYNT 840
 DB 781 KLVPFSSKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKRVLSTTLENRKENAYNT 840
 QY 841 SLSIIFSRNLHLASLTQPRSPPIKVECAAPSAHARLCSVGHVPVQTKAKVTFLEFEFSC 900
 DB 841 SLSIIFSRNLHLASLTQPRSPPIKVECAAPSAHARLCSVGHVPVQTKAKVTFLEFEFSC 900
 QY 901 SSLLSQVFGKLTASSDSLRNGTLQENTAQTSAYIOVEPHLLFSSSESTLHRYEVPYGTLL 960
 DB 901 SSLLSQVFGKLTASSDSLRNGTLQENTAQTSAYIOVEPHLLFSSSESTLHRYEVPYGTLL 960
 QY 961 PVGPGPEFKTTLR-----TNNASCIQVNL 985
 DB 961 PVGPGPEFKTTLRVQNLGCVVSGLIISALLPAVAHGNYFLSLSQVITNNASCIQVNL 1020
 QY 986 EPPGPPVHPELOHTNRLNGSNTQCQVVRCHLQALAGTEVSVGLLHVHNEFFRAKFK 1045
 DB 1021 EPPGPPVHPELOHTNRLNGSNTQCQVVRCHLQALAGTEVSVGLLHVHNEFFRAKFK 1080
 QY 1046 SLTVVSTFELGTREGSVLQLTEASRWSESLLEVVTQTRPILISLWILGVLGGLLALL 1105
 DB 1081 SLTVVSTFELGTREGSVLQLTEASRWSESLLEVVTQTRPILISLWILGVLGGLLALL 1140
 QY 1106 VFCLWKLGFPAH 1117
 DB 1141 VFCLWKLGFPAH 1152

RESULT 9
 AAB64658
 ID AAB64658 standard; protein; 1152 AA.
 XX AAB64658;
 AC AAB64658;
 XX 22-MAR-2001 (first entry)
 XX Human secreted protein BLAST search protein SEQ ID NO: 168.
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
 KW vulnervary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX Homo sapiens.

XX	PN	WO200077197-A1.	Qy	421	PPALQNHAAAYLYGYSVSMILLRGRRRLFLSGAPFRHRGKVIAFOLKKDQAVRVAQSLQGE	480
XX	PD	21-DEC-2000.	Db	421	PPALQNHAXYLYGYSXSKMLLRGKRLSGAXFRHRGKVIAFOLKKDQAVRVAQSLQGE	480
XX	PF	01-JUN-2000; 2000WO-US014934.	Qy	481	QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLO	540
XX	PR	11-JUN-1999; 99US-0138599P.	Db	481	QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLO	540
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	Qy	541	PEPPQDARPGFAMGALPDLNQDGFADVAVGAPLEDHGHQALYLYHGTQSGVRPHPAORIA	600
XX	PA	(ROSE/) ROSEN C A.	Db	541	PEPPQDARPGFAMGALPDLNQDGFADVAVGAPLEDHGHQALYLYHGTQSGVRPHPAORIA	600
XX	PI	Rosen CA, Ruben SM, Komatsoulis GA;	Qy	601	AASMPHALSYFGRSVDGRDLDDGDDLVDAVGAQGAAILSSRPVHLTPSLEVTPOAIS	660
XX	DR	WPI; 2001-032312/04.	Db	601	AASMPHALSYFGRSVDGRDLDDGDDLVDAVGAQGAAILSSRPVHLTPSLEVTPOAIS	660
XX	PT	Isolated nucleic acid molecule encoding a human secreted protein is used	Qy	661	VWORDCRRQOEAVCLTAALCFQVTSRTTCGRWPHQFYMRETASLDDEWTAGARAAFDGSGQ	720
XX	PT	in preventing, treating or ameliorating a medical condition.	Db	661	VWORDCRRQOEAVCLTAALCFQVTSRTTCGRWPHQFYMRETASLDDEWTAGARAAFDGSGQ	720
XX	PS	Disclosure; Page 547-551; 558pp; English.	Qy	721	RLSPRRLRLSVGNVTCBQLHFHVLDTSDYLRPVALTVPALDNTTKPGPVLNBSGSPSIQ	780
XX	CC	The invention relates to the isolation of genes AAP32757-F32803 encoding	Db	721	RLSPRRLRLSVGNVTCBQLHFHVLDTSDYLRPVALTVPALDNTTKPGPVLNBSGSPSIQ	780
XX	CC	the human secreted proteins AB64549-B64594. The sequence is used as a	Qy	781	KLVPFSDKCGPDNECVTDLVQVNMDIRSGRKAPFVVRGGRKVLVSTTLXKRKENAYNT	840
XX	CC	query sequence for doing BLASTX searches to identify homologous	Db	781	KLVPFSDKCGPDNECVTDLVQVNMDIRSGRKAPFVVRGGRKVLVSTTLXKRKENAYNT	840
XX	CC	sequences. The genes and proteins are useful for preventing, ameliorating	Qy	841	SLSLIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLLPEPFSC	900
XX	CC	or treating medical conditions, e.g. by protein or gene therapy. The	Db	841	SLSLIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLLPEPFSC	900
XX	CC	genes are isolated from a range of human tissues disclosed in the	Qy	901	SSLLSQVFGKLTASSDSLENGTLQNTAQTSAIOVEPHLLPSSBSTLHRYEVPYGTLL	960
XX	CC	specification. The nucleic acids, proteins, antibodies and (ant)agonists	Db	901	SSLLSQVFGKLTASSDSLENGTLQNTAQTSAIOVEPHLLPSSBSTLHRYEVPYGTLL	960
XX	CC	are useful in the diagnosis, treatment and prevention of: (a) cancer,	Qy	961	PVGPGPEFKTTLR-----TNNACIVQNLT	985
XX	CC	e.g. breast and ovarian cancer, and other cancers of the adrenal gland,	Db	961	PVGPGPEFKTTLR-----TNNACIVQNLT	985
XX	CC	bone, bone marrow, breast, gastrointestinal tract, liver, lung, or	Qy	986	EPGPGPVHPPELOHTNRLNGSNTCCQVVRCHLQQLAGTSEVSGLLRLVNEFPFRRAKPK	1045
XX	CC	urogenital; (b) immune disorders e.g. Addison's disease, allergies,	Db	986	EPGPGPVHPPELOHTNRLNGSNTCCQVVRCHLQQLAGTSEVSGLLRLVNEFPFRRAKPK	1045
XX	CC	autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,	Qy	1021	EPGPGPVHPPELOHTNRLNGSNTCCQVVRCHLQQLAGTSEVSGLLRLVNEFPFRRAKPK	1080
XX	CC	Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative	Db	1021	EPGPGPVHPPELOHTNRLNGSNTCCQVVRCHLQQLAGTSEVSGLLRLVNEFPFRRAKPK	1080
XX	CC	colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)	Qy	1046	SLTVVSTFELGTBEGSVLQTEASRWSESLLEVQVTRPILISLWILGSLGGLLLALL	1105
XX	CC	wound healing; (e) neurological diseases e.g. cerebral anoxia and	Db	1046	SLTVVSTFELGTBEGSVLQTEASRWSESLLEVQVTRPILISLWILGSLGGLLLALL	1105
XX	CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	Qy	1081	SLTVVSTFELGTBEGSVLQTEASRWSESLLEVQVTRPILISLWILGSLGGLLLALL	1140
XX	CC	and parasitic infections	Db	1081	SLTVVSTFELGTBEGSVLQTEASRWSESLLEVQVTRPILISLWILGSLGGLLLALL	1140
XX	Sequence 1152 AA;		Qy	1106	VFCLWKLGPFAH 1117	
XX			Db	1141	VFCLWKLGPFAH 1152	
XX	Query Match	97.2%; Score 5763.5; DB 4; Length 1152;	RESULT 10			
XX	Best Local Similarity	95.9%; Pred. No. 0;	ADR41424			
XX	Matches 1105; Conservative	3; Mismatches 9; Indels 35; Gaps 1;	ID	ADR41424	standard; protein; 1049 AA.	
Qy	1	MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPGPPEAEFGYSVLQHVGGQRWMLVGA	XX	AC	ADR41424;	
Db	1	VELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPGPPEAEFGYSVLQHVGGQRWMLVGA	XX	DT	07-OCT-2004 (first entry)	
Qy	61	PWDGPGDRRGDVYRCVPGGAHNAPCAKGLGDYQLGNSHHPAVNMHLGMSLLETDDGG	XX	DE	Human CD-like molecule HA0AD02, SEQ ID NO:223.	
Db	61	PWDGPGDRRGDVYRCVPGGAHNAPCAKGLGDYQLGNSHHPAVNMHLGMSLLETDDGG	XX	KW	Human; CD-like molecule; cluster of differentiation; diagnosis;	
Qy	121	FMACAPLWSRACGSSVSSGTCARVDASFPQGS LAPTAQRCTPYMDVIVLDGNSIYP	KW	KW	prevention; immune disorder; immunodeficiency; autoimmune disorder;	
Db	121	FMACAPLWSRACGSSVSSGTCARVDASFPQGS LAPTAQRCTPYMDVIVLDGNSIYP	KW	KW	blood-related disorder; haematological disorder; haemostatic disorder;	
Qy	181	WSEVQTFLRLVGLKFLTDPEQIQGLVQYGESPVHWSLGDFTKKEEVRAAKNLSRREG	KW	KW	thrombotic disorder; hyperproliferative disorder; cancer; tumour;	
Db	181	WSEVQTFLRLVGLKFLTDPEQIQGLVQYGESPVHWSLGDFTKKEEVRAAKNLSRREG	KW	KW	apoptotic disorder; cardiovascular disorder; respiratory disorder;	
Qy	241	RETTQAQIMVACTEGFSQSHGGRPEAALLVVVTDGSHDGBELPAALKACEAGRVTRY	KW	KW	angiogenic disorder; neovascularisation; neurological disorder;	
Db	241	RETTQAQIMVACTEGFSQSHGGRPEAALLVVVTDGSHDGBELPAALKACEAGRVTRY	KW	KW	endocrine disorder; reproductive system disorder; infectious disease;	
Qy	301	GIAVLGHLRQRDPSSFLREIRTIASDPDRPFNFVNTDEAALTDIVDALGDRIFGLEGS	KW	KW	gastrointestinal disorder; drug screening; tissue regeneration;	
Db	301	GIAVLGHLRQRDPSSFLREIRTIASDPDRPFNFVNTDEAALTDIVDALGDRIFGLEGS	KW	KW	chemotaxis; gene therapy; antibody therapy; drug targeting;	
Qy	361	HAENESSFGLBMSOIGSTHRLKDGILFGMVGYDWDGGSVLWLEGHRLFPFRMALEDEF	KW	KW	chromosome mapping; forensic analysis; immunophenotyping; cytostatic;	
Db	361	HAENESSFGLBMSOIGSTHRLKDGILFGMVGYDWDGGSVLWLEGHRLFPFRMALEDEF				

KW haemostatic; tranquiliser; vulnary; antiinflammatory; nephrotropic;
 KW cardiac; aniallergic; anti-HIV; antirheumatic; antiarthritic;
 KW antiporiatic; immunosuppressive; vasotropic; nootropic; neuroprotective;
 KW antithyroid; thyromimetic; gynaecological; virucide; hepatotropic;
 KW antibacterial; dermatological; chromosome 1q21.

XX Homo sapiens.

XX WO200226930-A2.

XX 04-APR-2002.

XX 25-SEP-2001; 2001WO-US029838.

XX 26-SEP-2000; 2000US-0235484P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Birse CE;

XX WPI; 2002-405050/43.

XX N-PSDB; ADR41248.

XX Novel polynucleotides and polypeptides useful for treating, preventing or
 PT ameliorating cardiovascular, renal, neurovascular, and autoimmune
 PT disorders.

XX Claim 11; SEQ ID NO 223; 1243pp; English.

XX The invention relates to 167 novel human CD (cluster of differentiation)-
 CC like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid:11)-

XX Sequence 1049 AA;

Query Match 82.2%; Score 4878.5; DB 5; Length 1049;

Best Local Similarity 91.1%; Pred No. 0;

Matches 955; Conservative 7; Mismatches 31; Indels 55; Gaps 4;

Qy 124 CAPLWSRA-CGSSVFSSGICARVDASFQPGSLAPTAQRCPTYMDV---VIVLDGNSIY 179

Db 18 CLQFDMWPCGCFIPASG-----KPGTHCPTLPNIHGCHCLGWLHSIY 61

Qy 180 PWSVQTFRLRLVCKLFDIPQIQVGLVQGESPVHWSLGDFTKSEVYVAAKNLSRRE 239

Db 62 PWSVQTFRLRLVCKLFDIPQIQVGLVQGESPVHWSLGDFTKSEVYVAAKNLSRRE 121

Qy 240 GRETKTAQAINVACTEGFSQSHGRPEAARLLVVVTDGESHGDBELPAALKACBAGRVR 299

Db 122 GRETKTAQAINVACTEGFSQSHGRPEAARLLVVVTDGESHGDBELPAALKACBAGRVR 181

Qy 300 YGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNNVTDBAALTDIVDALGDRIFGLEG 359

Db 182 YGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNNVTDBAALTDIVDALGDRIFGLEG 241

Qy 360 SHAENESSFGLMSQIGFSTHRLKDGILFGMVGYDWCSSVLWLEGGHRLPPRMALEDE 419

Db 242 SHAENESSFGLMSQIGFSTHRLKDGILFGMVGYDWCSSVLWLEGGHRLPPRMALEDE 301

Qy 420 FPPALQNHAAVLYGSVSSMLRGRRFLSGAPFRHRGKVIATFOLKKGAVRVAQSLQG 479

Db 302 FPPALQNHAAVLYGSVSSMLRGRRFLSGAPFRHRGKVIATFOLKKGAVRVAQSLQG 361

Qy 480 EQIGSYFGSELCPDTRDGTDTLLVAAPMFLGPQNKETGRVYVYLVGQOSLLTLQGTLL 539

Db 362 EQIGSYFGSELCPDTRDGTDTLLVAAPMFLGPQNKETGRVYVYLVGQOSLLTLQGTLL 421

Qy 540 QPEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTQSGVRPHPAQRI 599

Db 422 QPEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTQSGVRPHPAQRI 481

Qy 600 AAASMPHALSYFGSRVGRRLDGDGLVDVAVGAGGAAILLSRPIVHLTPSLEVTPOAI 659

Db 482 AAASMPHALSYFGSRVGRRLDGDGLVDVAVGAGGAAILLSRPIVHLTPSLEVTPOAI 541

Qy 660 SVVQDCCRGRQBAVCLTAALCFQVTSRTSGRMDHQYMRFTASLDQWTAGARAFDGS 719
 Db 542 SVVQDCCRGRQBAVCLTAALCFQVTSRTSGRMDHQYMRFTASLDQWTAGARAFDGS 601
 Qy 720 QRLSPRELRLSVGNVTCQQLHFHVDLSDYLRPVALTVPALDNTTKPGPVNLNGSPSTSI 779
 Db 602 QRLSPRELRLSVGNVTCQQLHFHVDLSDYLRPVALTVPALDNTTKPGPVNLNGSPSTSI 661
 Qy 780 QKLVPFSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGRRKVLVSTTLNKRKNAYN 839
 Db 662 QKLVPFSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGRRKVLVSTTLNKRKNAYN 721
 Qy 840 TSLSLIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPQTGAKVTFLEPEFS 899
 Db 722 TSLSLIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPQTGAKVTFLEPEFS 781
 Qy 900 CSSLLSQVFGKLTASSDLSLRNGTLQNTAQTSAIIOYEPHLLFSSSTLHRYEVHPYGT 959
 Db 782 CSSLLSQVFGKLTASSDLSLRNGTLQNTAQTSAIIOYEPHLLFSSSTLHRYEVHPYGT 841
 Qy 960 LPVGPGEPEFKTTLR-----TNNASCIQVNL 984
 Db 842 LPVGPGEPEFKTTLRVQNLGCVVWSGLIISALLPAVAGHGNVFLSLSQVITNNASCIQVNL 901
 Qy 985 TEPGPPVHPPELOHTNRLNGSNTQCVVRCHLQKAGTEVSVGLRLVHNEFRRAKF 1044
 Db 902 TEPGPPVHPPELOHTNRLNGSNTQCVVRCHLQKAGTEVSVGLRLVHNEFRRAKF 961
 Qy 1045 KSLTVSTFELGTEEGSVLQLTASRWSSESLVWVQTRPILISLWILIGSVLGGLLALL 1104
 Db 962 KSLTVSTFELGTEEGSVLQLTASRWSSESLVWVQTRPILISLWILIGSVLGGLLALL 1021
 Qy 1105 LVFCLWKLGFFAHKKIPEEEKREKLEQ 1132
 Db 1022 LVFCLWKLGFFAHKKIPEEEKREKLEQ 1049
 RESULT 11
 AAU14231
 ID AAU14231 standard; protein; 1188 AA.
 AC AAU14231;
 XX AAU14231;
 DT 24-OCT-2001 (first entry)
 XX Human novel protein #102.
 DE Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX Homo sapiens.
 OS OS
 XX WO200155437-A2.
 XX 02-AUG-2001.
 PD 25-JAN-2001; 2001WO-US002623.
 PF 25-JAN-2000; 2000US-00491404.
 PR (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-451939/48.
 XX N-PSDB; AAS22536.

Isolated polypeptides useful for treating anti-inflammatory diseases, nervous system disorders, and for regenerating bone and cartilage.

Example 4; Page 578-581; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. CC Polynucleotides of the invention are used as probes and primers, for sequencing, for chromosome or gene mapping, in the production of recombinant proteins, and in generating anti-sense DNA or RNA and in gene therapy. Polypeptides of the invention can be used to target drugs to a tumour, in assays to determine biological activity, to raise antibodies/ elicit an immune response, to determine quantitative protein levels, as tissue markers, and to isolate receptors or ligands. CC Polypeptides of the invention may also be useful in treating platelet disorders, stem cell disorders, regenerating bone, cartilage, tendon, ligament and/or nerve tissue, wound healing, treating burns, promoting the proliferation, differentiation and survival of stem cells, as a contraceptive, treating osteoporosis and osteoarthritis, anaemia, Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, stroke, immune deficiencies resulting from bacterial, viral or fungal infection or from autoimmunity, cancer, allergy, asthma, graft-versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory diseases, nervous system disorders, and infection. The present sequence CC represents a protein of the invention

Sequence 1188 AA;

Query Match 39.7%; Score 2355; DB 4; Length 1188;
Best Local Similarity 42.3%; Pred. No. 3.38-201;
Matches 506; Conservative 204; Mismatches 406; Indels 80; Gaps 16;
QY 1 MELPFVTHFLVFLVFLGLCSPPNLDHHPFLPGPPEAFEGYSVLQHVGGQRWMLVGA 60
DB 1 MDLPRLGVVAWALSMLPGFTDTFMDTRKPRVPGSRTPFGTYTVQQHDISGNKMLVGA 60
QY 61 PWDPGSDRGDDVRCVPGVGHNAAPCAKHGLDYLQNSHPVAMHGLMSLLTDDGG 120
DB 61 PLETNGYQKTGDYKCPV---IHGCTKLNGLRVTLNVSERKONMRLGLSLATNPKDNS 117
QY 121 FMACAPLWSRACGSSVFSSGTCARVDASFQPGQSLAPTACRCPTYMDVIVLDGNSIYP 180
DB 118 FLACPLWSHECGSSYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
QY 181 WSEVQTLRLRLVGLFIDPEIQVLQVQGESPVHWSLGDFTKKEEVRAAKNLSRREG 240
DB 178 WVEVQHFLINLKKFYIGPGQIQGVVQYGEDVVEHFLNDYRSVKDVEAASHIEQGG 237
QY 241 RETTQAQIMVACTEGFSQSHGGPEARLLVVVTDSGHDGBELPALKACAGRVTRY 300
DB 238 TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOISERDNTVRY 295
QY 301 GIAVLGHVLRQRDPSSFLREITASDPDERFFNVTDDEAALTDIVDGLDRIPLGLEGS 360
DB 296 AVAVLGYNNRGINPEFLNEIKYIASDDPKHFNVTDEAALDKDIVDALGDRIFSLSEGT 355
QY 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGAYDWGGSVLWLGSHRLPFPMALEDEF 420
DB 356 N-KNETSFGLEMSQTGSSHVVVEDGVLLGAVGAYDMNGAVLKETSGAKVIPLESYLKEF 414
QY 421 PPALQNHAAVLYGYSSVSMMLRGRRLLFSGAPRFRHKGKVIJAFOLKKGAVRVAAQSLQGE 480
DB 415 PEELKNHGAYLYGTVTVTSVSSRQGRVYVAGAPRFNHTGKVLFTMHNRSLLTHQAMRGQ 474
QY 481 QIGSYFSGSELCPDTRDGTDDVLLVAAPMLFGPNKGTGRVYVYVQVQSLTLTQTLQ 540
DB 475 QIGSYFSGSEITSDIGDGVTDVLLVGAPMTFN--EGREGKVVYIEL--RQNRFYVNGTLK 532

QY 541 PEPP-QDARPGFAMGALPDLNQDGFADVAVGAPLEDHQGALYLYHGTQSGVRRPAPQRI 599
DB 533 DSHSYQNARFGSSIASVRLNQDSYNDVVVVGAPLEDNHAGAIYIFHGRGSIILKTPQRI 592
QY 600 AAASMPHALSYFGSRVDGLDDGLVDVAVGAQGAAILSSRPVHLTPSLVETPQAI 659
DB 593 TASELATGQYFGCSIHGQJDLNEDGLIDVAGALGNVILWSPVQVQINASHFEPSKI 652
QY 660 SVVQDRCCRRRQBAVCITAAALCFQVTSRTSGRWDHGFYMRFTASLDEWTAGAAAFDGS 719
DB 653 NIFHRDCKSRGRATCLAAFLCTPFLAPHFTQTVTVGIRYNATMDERRYPRAHLEGG 712
QY 720 QRLSPRLRLSVGNVTCQLHFVLDTSYLRVALVTVPALDNTTKPGVILNMGSTSI 779
DB 713 DRFTNRAVLSSQELCERINFHLVDADYVKVPTFSVEYSLEDPDH-GPMLDDGWPTTL 771
QY 780 QKLVPSKDCGPNCECTDVLQVNDI-----RGRK-----APFVV 817
DB 772 RVSVPPFNGCNEDEHCVLDJLDARSDLPTAMBYCQRLKPAQDCSAYTSLSDTTVFII 831
QY 818 RGRKRVLVSTTLNKRKENAYNTSLSIIFSRNLHLASLTPQRSPIKVECAAPS--AHAR 875
DB 832 ESTQRQVAVETLENGENAYSTVLNISQSANLPASLIQKEDSDGIECWNERLQKQ 891
QY 876 LCSVGHVPFGTGAQVTLLEFEPSCSLLSQVFGKLTASSDSLSERNGTLOENTAQTSAYI 935
DB 892 VCVSYPPFRAKAKVAPRLDFEFPSKIFLHHLBIELAAGSDSNERDSTKEDNVAPLRFHL 951
QY 936 QYEPHLLPSESILHRYEYHPYCTLP--VCPGPEFKTLRTN----- 975
DB 952 KYEADVLFTSRSSLSHYEVKPNSSLERYDYGIPFPFCIFRIQNLGLFPIHMMMKITPI 1011
QY 976 -----NASC-IVQNLTEPPGPPVHPEELOHTNRLNGSNTCCQVVR 1015
DB 1012 ATRSGNRLKLRLDPLTDEANTSCNIWGNSTETPTVE-EDLRAPOLNHSNDVVSINC 1070
QY 1016 HLOQLAKGTVEVGLRLVHNEFFRAKFKSLTVVTFELGTBEGSVLQVTEASRWSES 1075
DB 1071 NI-RLVNPQIEINFLGNLWRLSLKALKYKSMKIMVNAALQRPSPFIPREEDPSQIV 1129
QY 1076 LEVQTRPILISLWILIGSVLGGILLALVFLCKLGFPAHKKIPPEEKREKLE 1131
DB 1130 FEISKQEDWQVPIWIVGSTGLGLLLALLVLAALWKLGFPR SAR-----RRREPGLD 1181
RESULT 12
AAB50085
ID AAB50085 standard; protein; 1188 AA.
XX
AC AAB50085;
XX AC
XX 19-MAR-2001 (first entry)
DT
XX Human A259.
DE
XX Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
KW rheumatoid arthritis.
XX
OS Homo sapiens.
XX
FH Key
FT Domain
FT /label= Extracellular_domain
FT Peptide
FT /label= Signal_peptide
FT Protein
FT /label= Mature_protein
FT Domain
FT /label= Integrin_alpha_subunit_repeat_domain_#1
FT Domain
FT /label= Integrin_alpha_subunit_repeat_domain_#2
FT Domain
FT 164..345

```

FT FT Domain /label= I domain
FT FT 367..392_
FT FT /label= integrin_alphasubunit_repeat_domain_#3
FT FT 421..455
FT FT /label= Integrin_alphasubunit_repeat_domain_#4
FT FT 478..516
FT FT /label= Integrin_alphasubunit_repeat_domain_#5
FT FT 540..575
FT FT /label= Integrin_alphasubunit_repeat_domain_#6
FT FT 602..640
FT FT /label= Integrin_alphasubunit_repeat_domain_#7
FT FT 1142..1164
FT FT /label= Transmembrane_domain
FT FT 1165..1188
FT FT /label= Cytoplasmic_domain
XX XX W0200073339-A1.
XX XX
XX XX 07-DEC-2000.
XX XX
XX XX 15-MAY-2000; 2000WO-US013262.
XX XX
XX XX 28-MAY-1999; 99US-00322790.
XX XX 27-APR-2000; 2000US-00561263.
XX XX
XX XX (MILL-) MILLENNIUM PHARM INC.
XX XX
XX XX Pan Y, Lora JM;
XX XX
XX XX WPI; 2001-041142/05.
XX XX
XX XX N-PSDB; AAC91901, AAC91902.
XX XX
XX XX Nucleic acid encoding alpha-integrin subunits, useful for treatment and
XX XX diagnosis of fibrosis, e.g. of the liver.
XX XX
XX XX Claim 8; Fig 1; 164pp; English.
XX XX
XX XX The present sequence is human integrin alpha subunit, A259. A259 is
XX XX homologous with the alpha1 and alpha10 integrin subunits and is
XX XX overexpressed in fibrosis. A259 is implicated in regulation of
XX XX proliferation, differentiation and/or function of many different cell
XX XX types. Inhibitors of A259 activity are useful for the treatment of liver
XX XX disease, particularly fibrosis, and also fibrosis in other organs
XX XX (specifically lung and kidney). In addition, A259 can be used for
XX XX treatment and prevention of cancer, osteoporosis, acute myeloid
XX XX leukaemia, HIV infection, and rheumatoid arthritis
XX XX
XX XX Sequence 1188 AA;

Query Match 39.7%; Score 2355; DB 4; Length 1188;
Best Local Similarity 42.3%; Pred. No. 3.3e-201;
Matches 506; Conservative 204; Mismatches 406; Indels 80; Gaps 16;

Qy 1 MELPFTVHLFLVFLVTLGLCSPENLDEHHPLRFPQPEAEFGYSVLQHVGGQRMVLGA 60
Db 1 MDLPRLGVVAWALSILWPGFTTFNMDTKKRVIPGSRATFAGYTVQQHDISGNKWLWVGA 60
Qy 61 PWDGPGDRRGDGYRCPVGGAHNAPCAKHGLGDIQLGNSSHPAVNMHGLMSLLETGDDGG 120
Db 61 PLETNGYQKTDVYKCPV---IHGNCYTKNLGRVTLNSVSEKDNMRGLSLATNPKDNS 117
Qy 121 FMACAPLWSRACGSGSFSSGICARVDASFPQGSGLAPTAQRCPYMDVWVILVLDGNSIYP 180
Db 118 FLACSPLWSHCEGSSYVTGMCVRNSNFRPSKTVAPALQRCQTYMDIVILVLDGNSIYP 177
Qy 181 WSEVQTLRLRLVGLFDPEIQVGLQVYGSPVHWSLGRFRTKEEVVRAAKNLSRREG 240
Db 178 WVEVQHLINLTKFYIGPGQIQGVQYQGVGDVWHFHLNDYRSVKDVEAAHIEQRGG 237
Qy 241 RETTAAQIAIMVACTEGFSQSQSHGRPEAARLLVWVTVDGSHDGEELPAALKACAEAGRVTRY 300
Db 238 TETRTAGIEPAREAQK--GGRKGAKKVMIVITDGESHSDPDLKVKYIQOSERDNTVTRY 295

301 GIAVLGHVLRQRDPSSFLREIRTIASDPDERPFNFVNTDEAALTDIYDALGDRIFGLGGS 360
326 AVAVLGYNNRRGINPETFLNEIKYIADPDOKHFNFTVDEAALKDIYDALGDRIFSLDEGT 355
361 HAENESSFGLSEMSQIGFSTHRLKDGILFGMVGADWGGSVLWLEGGHRLFPFRMALEDEF 420
386 N-KNETSFGLSEMSQTFSSHVVEDGVLGAVGADWNGAVLKETSAGKVIPLRESYLKEF 414
421 PPALQNHAAVLYGSVSNMLLRGGRLFLSGAPRFRHKGKVIAPOLKDKGAVRVQAQSGE 480
445 PEELKNGHAYLGYTVTSVSSRQGRVTVAGAPRHNHTGKVLFTMHNRRSLTIHQAMRGQ 474
481 QIGSYFSGELCPDTRDGTDLVLAAPMLFGPQNKETGRVYVLYVQCSLLTLTGTLQ 540
495 QIGSYFSGEITSVDIDGDTVLLVAGPMYFN-EGRERGKVVYVEL-RQRFRVYVNGTLK 532
541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDHGGALYLYHGTSQSVRPHPAQRI 599
566 DSHSYQNARFGSSIASVRDLNQDSYNDVVVGALEDNHAGAIYIFHGFSGILKTPKQRI 592
600 AAASMPHALSYFGRSDVGRDLDDGDLVDVAVAGGAAIILSSRPIVHLTPSLEVTPQAI 659
625 TASELATCLOVFGCSIHGQLDLNEDGLIDLAVGALGNVILWSRPVQINASLHPEPSKI 652
660 SVVORDCRRRGOEAVCLTAALCFQVTSRTSGRWDHQFYMREFASILDDEWTAGARAFDCSG 719
685 NIFHRDCKRGRDATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYTPRAHLDEGG 712
720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTTFALDNTTKPGPVLNBSPTS 779
745 DRFTNRAVLJSSGOELCERINFHVLDTADYVKPVTFSVEYSLEDPDH-GPMLDDGWPTTL 771
780 QKLVFPGKDCPDNECVTDVLQVNMDI-----RGRK-----APFV 817
805 RVSPFWMGNCNEDHCVPDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTSLSFDTTVEII 831
818 RGRKRVLTTLNRRKENAYNTSLIFSNLHLASITPQRESPIKVECAAPS--AHAR 875
843 ESTQRVAVEATLENGENAYSTVLNISQSANLQFASLIQKEDSDGSECYNEERRLOKQ 891
876 LCSVGHVPVOTGAKVTFLLBEPFSCSLLSOVFKLTASSDSLERNGLQNTAQTSAI 935
892 VCNVSYFFRAKAKAVRFLDFEFSKSIPLHLELELAAGSDSNERNSTKEDNVAPLRHL 951
936 QYEPHLLFSSESTLHYRVEHPYGLP--VGPGEPEKTTLRN----- 975
952 KYEADVLFTRSSLSLHYEVKPNSSLERYDVGIGPPFCIFRIONLGLPFIHGMWKITPI 1011
976 -----NASC-IVQNLTEPPGPPVHPELOHTNRLNGSNTQCQVRC 1015
1012 ATRSGNRLKLRLDFTLDEANTSCNIGNSTREYRTPVE-EDLRAPQLNHSNSDVVSINC 1070
1016 HLGLAKGTBVSVGLLRLVHNEPFRRAKFSLTVVSTFELGTBEGSVLQLTEASRWSL 1075
1071 NI-RLVFNQEIHFLLGNLWRLSKALKYKSMKIMVNAALQRPSPPIFREEDPSRIV 1129
1076 LEVQTPRILISLWILGSLVGLLLALLVFLCWLKLGFFFAHKKIPBEEKREEKLE 1131
1130 FEISKQSDWQVPIWIVGSTLGLGLLLALLVLAWLKLGFFFSAR----RRREPGLD 1181

RESULT 13
AAU10551
ID AAU10551 standard; protein; 1188 AA.
XX
XX AAU10551;
AC
AC 14-FEB-2002 (first entry)
DT
DT Human A259 polypeptide.
DE
DE Human; A259; integrin alpha subunit; integrin alpha 10; secreted protein;
KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;

```

cartilage associated disorder; haematopoietic disorder; bone marrow; immune related disease; apoptotic disorder; neuronal tissue disease; neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic; nephrotropic; immunomodulator; anti-inflammatory; neuroprotective; antiarthritic; antianemic; antiallergic; antiaschmatic; dermatological; antidiabetic; anticonvulsant; antiparkinsonian.
 Homo sapiens.
 Key Location/Qualifiers
 Domain 1..1141
 Peptide 1..22
 Protein 23..1188
 Domain 37..90
 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
 Domain 115..157
 /note= "Integrin alpha repeat domain"
 Domain 164..345
 /note= "I domain or Von Willebrand Factor type A domain"
 Domain 367..392
 /note= "Integrin alpha repeat domain"
 Domain 421..472
 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
 Domain 476..532
 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
 Domain 538..593
 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
 Domain 600..654
 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
 Domain 1142..1164
 /note= "Transmembrane domain"
 Domain 1165..1188
 /note= "Cytoplasmic domain"
 WO200181414-A2.
 01-NOV-2001.
 27-APR-2001; 2001WO-US013516.
 27-APR-2000; 2000US-00561263.
 (MILL-) MILLENNIUM PHARM INC.
 Pan Y, Lora J;
 WPI; 2002-041397/05.
 N-PSDB; AAS16873.
 New A259 nucleic acids and polypeptides, which comprise integrin alpha subunit, useful for diagnosing, preventing or treating e.g. liver disease, kidney or lung fibrosis, cancers, blood disorders or immune related diseases.
 Claim 9; Fig 1; 168pp; English.
 The invention relates to human and murine A259 nucleic acid molecules which encode secreted proteins with homology to integrin alpha subunits, specifically to integrin alpha 10. The A259 polypeptide and nucleic acid are useful for treating liver disease or fibrosis, particularly kidney fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also useful for diagnosing, preventing or treating cartilage and bone associated disorders (such as bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and osteoporosis), bone marrow, blood and haematopoietic disorders (such as

CC acute myeloid leukaemia, haemophilia, anaemia and thalassaemia), immune related diseases (such as HIV, viral infections, cancers, T cell autoimmunity disorders e.g. AIDS, and allergic inflammatory disorders e.g. asthma and psoriasis), apoptotic disorders (such as systemic lupus erythematosus and insulin-dependent diabetes mellitus), diseases of the neuronal tissues (such as epilepsy and muscular dystrophy) and neurodegenerative diseases (such as Parkinson's disease and Huntington's disease). This sequence represents the human A259 polypeptide
 XX
 SQ Sequence 1188 AA;
 Query Match 39.7%; Score 2355; DB 5; Length 1188;
 Best Local Similarity 42.3%; Pred. No. 3.3e-201;
 Matches 506; Conservative 204; Mismatches 406; Indels 80; Gaps 16;
 QY 1 MELPFTVHLPLVFLVLTGLCSPEFLPFPPEAFEGYSVLQHVGGGQWMLVGA 60
 DB 1 MDLPRLGLVAVMALUSLWPGFTDTFMDTRKPRVLPGRRTAFYVQVQHDLSGNKWLTVGA 60
 QY 61 PWDGPGSDRRGDVYRCPVGGAHNAPCAKGLHGDYQLGNSHPAVNMHLGMSLLETDDGG 120
 DB 61 PLSTNGYQKTGDVYKCPV---IHGNCCTKLNLRVTLNSVSEKDNMRGLSLATNPKN 117
 QY 121 FMACAPLWSPACSSVFSFGICARVDASFOQSGSLAPTAQRCPTYMDVIVLVDGNSIYP 180
 DB 118 FLACSPILMSHECGSSYTTGMCRSVNSFRFTKVPALQRCQTYMDIVLVDGNSIYP 177
 QY 181 WSEVQVFLRLVGLKFLDPEQIQVGLVQYGESPVHEWSLGDFTKEBVRAAKNLSREG 240
 DB 178 WVEVQVFLRLVGLKFLDPEQIQVGLVQYGESPVHEWSLGDFTKEBVRAAKNLSREG 237
 QY 241 RETKTAQIMVACTEGFSQSHGGRPEAARLLVVVTVDGSHDGBELPAALKACAGRVTRY 300
 DB 238 TETRTAFGIETAFSEAFQK--GGKGAKKVMIVITDGHSDSPDLKVIQSSERDNTVRY 295
 QY 301 GIAVLGHLRQRDPSFLREITIASDPPERFPFNTDEAALTIDVADLGRDIFGLEGS 360
 DB 296 AVAVLGYNRRGINPETFLNEIKYIASDDPKHFPFNTDEAALKDIYDALGDRIFSLEGT 355
 QY 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDVGSGVLMLEGGHRLPPPMALDEDF 420
 DB 356 N-KNETSFGLMSQIGFSTHRLKDGILFGMVGYDVGSGVLMLEGGHRLPPPMALDEDF 414
 QY 421 PPALQHAAYLGYSVSMILRGRRRLFLSCAPFRHRGKVIATFQKKDGAVRVAQSLQGE 480
 DB 415 PEELKNHGYALGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSITIHQMRGQ 474
 QY 481 QIGSYFGSELCPIDTRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTLLQ 540
 DB 475 QIGSYFGSEITSDIDGDGVTDLVLLVAGAPMYFN--EGRERGVYVYEL--QNRNFYNGTLK 532
 QY 541 PEPP-QDARFGFANGALPDLNODGPADVAVGAPLEDGHGALVLYHGTQSGVRPHPAQRI 599
 DB 533 DSHSYQNAIRFGSSIASVRDLNQSDYNDVVVVGAPLEDNHAGATYIFHGRSILKTPQRI 592
 QY 600 AAASMPHALSYFGSRVDGRLDDGLDLDVAVAGAGAAIILSSRPVHLTPSLEVTQAI 659
 DB 593 TASELATGLQYFGCSIHGQLDLEDGLDLAVGALGNVILWSRPVQINASLHFEFSKI 652
 QY 660 SVVQDRCRRGQBAVCLTAALCFQVTSRTGRWDHQYMPRTASLDEWTAGARAAPFGSG 719
 DB 653 NIFHRDCKSRGRDATCLAAFLCFTPIFLAPHFQTTVIGIRYNATMDRERYTPRAHLDEGG 712
 QY 720 QRLSPRLRLSVGNVTCTEQLHFHVLDTSDYLRPVALTFTFALDNTTKPGPVNLGSGPTSI 779
 DB 713 DRFTNRAVLVLSGQELCERINFHVLDTADYVKKVPTFSVEYSLEDDPH--GPMLDGWNFTLL 771
 QY 780 QKLVPFSGKDGPNCEVDTDLVLQVNMDI-----RGSRK-----APFVV 817
 DB 772 RVSVFPWNGCNEDEHCVPDLVLDARSLLPTAMEYQVRVLRKPAQDCSAYTSLPDTTFII 831
 QY 818 RGRKRVLVSTTLNENKENVNTSLSIIFSRNLHLASLTTPQRESPIKVECAPS--AHAR 875
 DB 818 RGRKRVLVSTTLNENKENVNTSLSIIFSRNLHLASLTTPQRESPIKVECAPS--AHAR 875

Db 832 ESTQRVAVATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIETCVNEERRLQKQ 891
 Qy 876 LCSVCHVPFGCAKVTFLLEFEFSCSLLSQVFGKLTASSLSLRNGTLQNTAQTSAVI 935
 Db 892 VCNVSYPFRAKAVAFRLDFEFSKIFLHLELELAAGSDSNERDSTKEDNVAPLPHL 951
 Qy 936 QYEPHLLFSSBSTLHRYVHPVGTLP--VGQPEFKTTLRTN----- 975
 Db 952 KYEADVLFTRSSLSHYEVKNSSLYRDYDGLGPFSCIFRQNLGLFPIHGMKMTIPI 1011
 Qy 976 -----NASC-IVQNLTPPGPPVHPPELQHTNRLNGSNQCVVRC 1015
 Db 1012 ATRSGNRLKLLRDLTDEANTSCNIGNSTEYRPTVE-EDLRRAPQLNHSNDVWSINC 1070
 Qy 1016 HLGOLAKTEVSGVGLLRVHNEFFRRAKFSKLTVVSTFELTEGSGVQLTASRWBSL 1075
 Db 1071 NI-RLVPNQEIHFLLGNLWLRSLKALKYKMKMVNAALQRFHSPPIFEEDPFSRQIV 1129
 Qy 1076 LEVVQTRPILISLWLGSLVGLLLALLVFLWKLGFFAHKKIPBEEKREKLE 1131
 Db 1130 FEISKQEDWQPIWIIIVGSTLGGLLALLVALLWKLGGFFRSAR-----RRREPGLD 1181

RESULT 14

AAU14467
 ID AAU14467 standard; protein; 1188 AA.

XX AC AAU14467;

XX DT 24-OCT-2001 (first entry)

XX DE Human novel protein #338.

XX KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiaesthetic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.

XX OS Homo sapiens.

XX PN WO200155437-A2.

XX PD 02-AUG-2001.

XX PF 25-JAN-2001; 2001WO-US002623.

XX PR 25-JAN-2000; 2000US-00491404.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-451939/48.

XX RX N-PSDB; AAS22772.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,

XX nervous system disorders, and for regenerating bone and cartilage.

XX Example 4; Page 828-831; 894pp; English.

XX The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.

CC Polynucleotides of the invention are used as probes and primers, for
 CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene

CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise
 CC antibodies/elicit an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC represents a protein of the invention

XX SQ Sequence 1188 AA;

Query Match 39.7%; Score 2354; DB 4; Length 1188;

Best Local Similarity 42.3%; Pred. No. 4.1e-201;

Matches 506; Conservative 204; Mismatches 406; Indels 80; Gaps 16;

Qy 1 MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLPGPEAEFGYSLQHVGGQRMVLVGA 60
 Db 1 MDLPRGLVVAWALSLLWPGFTDTFNMDTRKPRVIFGSRATFQYTVQQHDISGNKVLVGA 60
 Qy 61 PMDGSGDRRGDYVRCVPGGAHNAAPCAKHGLDYLQNGNSHPAVNMHLMGMLLETDDGG 120
 Db 61 PLETNGYQKTGDVYKCPV---IHGNCITKLNLGRVTLNVSERKDNMRLLGLSLATNPKDNS 117
 Qy 121 FMACAPLWSRACSSSVFSSGICARVDASFQPGSLAPTAQRCPTMDVVILDCGNSLYP 180
 Db 118 FLACSPWSHECGSSYYTGMCSRVNSFRSKTVAPALORCQTYMDIVIVLDCGNSLYP 177
 Qy 181 WSEVQTELRRLVGLKFLIDPEIQVGLVQGESPVHMSLGDPRFRTKEEVRAAKNLSRREG 240
 Db 178 WVEVQHEFLINLLKFFIGPQIQVGVVQGEDVVEHFLNDYRSVKDVEAASHIEQREG 237
 Qy 241 RETKTAQIAIVACTEGFSQSHGGRPEARLLVVVTGESHGDEELPAALKAACEAGRVTRY 300
 Db 238 TETRTAFGIFARSEAFQK--GGRKGAQVMIVITDGESHSDSPDLEKVIQOOSRDNVTRY 295
 Qy 301 GIAVLGHVLRQRDPSSFLREIRTIASDPDERPFNVTDEAALTDIVDALGDRIFGLEG 360
 Db 296 AVAVLYGYNRRGINPETFLNEIKYIASDPDDPKGFNVTDEAALXDIVDALGDRIFSLEGT 355
 Qy 361 HAENESSFGLEMSQIGFSTHRLKDGILFGVMGAYDWGSLWLGSGHRLPPRMALEDEF 420
 Db 356 N-KNETSFGLEMSQTFSSHVVEDVLLGAVGAYDWGAVLKETSAGKVIPLRESYLKEF 414
 Qy 421 PPALQNHAAAYLGYSVSMLLRGGRRLLFSGAPRFRHGRKVIAPQLKKDGAVRVAQSLQGE 480
 Db 415 PEELKNHGAYLGYTTSVSSRQGRVVVAGAPRPNHTGKVLFTMNNRSLTIHQAMRGQ 474
 Qy 481 QIGSYFGSELCPDLTDRDGTDLVLLVAAPFLGPONKETGRVYVYLQVQGSLLTQGLTQ 540
 Db 475 QIGSYFGSEITSDVIDDGDGVTDVLLVGAPMYFN--EGREGRKVVYVEL--RQNRFYVNGTLK 532
 Qy 541 PEPP-QDARFGFAGALPDNLQDGFADVAVGAPLEDHOGALYLYHGTQSGVRPHPAORI 599
 Db 533 DSHSYQNRFGSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPKORI 592
 Qy 600 AAASMPHALSYFGSRVSDGRDLDDGLVDVAVAGQGAAILSSRPVHLTPSLEVTPQAI 659
 Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNAVILWSRPVQINASLHFPESKI 652
 Qy 660 SVVQDRCRRGQEAIVCLTALCFOVTSRTGFRWDHQYMRFTASLDEWTAGARAAFDGSG 719
 Db 653 NIFHRDCKRSGRDATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYPRAHLDEGG 712
 Qy 720 QRLSPRLRLSVGNVTCEQLHFEHLDTSDYLRLPVALTVTTFALDNTTKPGPVNLSGSPISI 779
 Db 713 DRFTNRAVLSSGQELCERINFHVLDTADYVKKPTVFSVEYSLEDPDH-GPMLDDGWPTTL 771

Db 832 ESTQRVAVATLENRGENAYSTVLNISQSANLOFASLIQKEDSDGSIECVNEERRLOKQ 891
Qy 876 LCSVGHVPVOTGAKVTFLLEFEFSCSSLLSQVFGKLTASSDSLERNCTLOENTAQTSAYI 935
Db 892 VCNVSYPPFFRAKAKVAFRLDPEFSKSIPLAHLEIETELAAGSDSNERDSTKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSSESTLHRYEVHPYGTLP--VGPGPEFKTTLRTN----- 975
Db 952 KYEVDVLFTRSSLSLHYEVKPNSSLERYDYGIGPPFCIFRIQNLGLPFIHGMMMKITIPi 1011
Qy 976 -----NASC-IVQNLTEPPGPPVHPEELOHTNRLNGSNTQCQVRC 1015
Db 1012 ATRSGNRLKLRDFLTDEANTSCNIWGNSTERYPTVE-EDLRRAPQNLHNSNDVVSINC 1070
Qy 1016 HLGQLAGTEVSVGLLRLVHNEFFRRAKFKSLTVVSTFELGTEGVSVLQLTEASRWSESL 1075
Db 1071 NI-RLVPNQEIHFLLGNLWLSLKALKYKSMKIMVNAALQRFHSPFIIFREEDPSRQIV 1129
Qy 1076 LEVVQTRPILISLWILIGSVLGGLLALAVFCLWKLGFPAHKKIPPEEKREEKLE 1131
Db 1130 FEISKQEDWQVPIWIVGSTLGGLLLALLVLALWKLGFPSAR-----RRREPGLD 1181

Search completed: April 6, 2005, 12:22:35
Job time : 149.162 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:17:21 ; Search time 141.389 Seconds

(without alignments)
2658.060 Million cell updates/sec

Title: US-09-647-544-4

Perfect score: 5932

Sequence: 1 MELPFVTHLFLPLVLTGLC.....GFAHKKIPBKKREKLEQ 1132

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 1418010

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep:*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubaa/US10D_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubaa/US11_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubaa/US11_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5890.5	99.3	1167	16	US-10-741-601-531 Sequence 531, App
2	5890.5	99.3	1177	16	US-10-741-601-532 Sequence 532, App
3	2355	39.7	1188	15	US-10-291-265-338 Sequence 338, App
4	2354	39.7	1188	15	US-10-291-265-810 Sequence 810, App
5	2344.5	39.5	1189	10	US-09-984-130-35 Sequence 35, Appl
6	2344.5	39.5	1189	10	US-09-836-353A-35 Sequence 35, Appl
7	2344.5	39.5	1189	15	US-10-262-839-4 Sequence 4, Appl
8	2194	37.0	1034	10	US-09-984-130-43 Sequence 43, Appl
9	2194	37.0	1034	10	US-09-836-353A-43 Sequence 43, Appl
10	2140	36.1	1120	15	US-10-262-839-6 Sequence 6, Appl
11	1897.5	31.8	1179	10	US-09-918-715-250 Sequence 250, App
12	1884	31.8	1180	10	US-09-918-715-307 Sequence 107, App
13	1881.5	31.7	1151	10	US-09-984-130-103 Sequence 103, App

14	1881.5	31.7	1151	10	US-09-836-353A-103	Sequence 103, App
15	1874.5	31.6	1177	17	US-10-741-600-1161	Sequence 1161, App
16	1733.5	29.2	1181	14	US-10-160-354-2	Sequence 2, Appl
17	1733.5	29.2	1181	15	US-10-295-027-1286	Sequence 1286, App
18	1733.5	29.2	1181	15	US-10-211-462-187	Sequence 187, App
19	1727.5	29.1	707	9	US-09-764-870-313	Sequence 313, App
20	1727.5	29.1	707	14	US-10-125-540-313	Sequence 313, App
21	1719.5	29.0	1147	15	US-10-336-603A-42	Sequence 42, Appl
22	1718	29.0	1148	16	US-10-872-198-147	Sequence 147, App
23	1126	19.0	1161	10	US-09-350-259-2	Sequence 2, Appl
24	1124	18.9	1161	9	US-09-891-943-2	Sequence 2, Appl
25	1124	18.9	1161	9	US-09-350-259-55	Sequence 55, Appl
26	1124	18.9	1161	10	US-09-891-943-55	Sequence 55, Appl
27	1120.5	18.9	1161	9	US-09-350-259-99	Sequence 99, Appl
28	1120.5	18.9	1161	10	US-09-891-943-99	Sequence 99, Appl
29	1116.5	18.8	1170	17	US-10-741-600-1088	Sequence 1088, App
30	1115.5	18.8	1170	9	US-09-945-265-2	Sequence 1, Appl
31	1114.5	18.8	1170	15	US-10-261-164-1	Sequence 1, Appl
32	1114.5	18.8	1161	9	US-09-350-259-53	Sequence 53, Appl
33	1114.5	18.8	1161	10	US-09-891-943-53	Sequence 53, Appl
34	1111	18.7	1145	16	US-10-872-198-130	Sequence 130, App
35	1110	18.7	1223	16	US-10-408-765A-295	Sequence 295, App
36	1110	18.7	1223	17	US-10-741-600-1086	Sequence 1086, App
37	1108.5	18.7	1155	9	US-09-350-259-46	Sequence 46, Appl
38	1108.5	18.7	1155	10	US-09-891-943-46	Sequence 46, Appl
39	1108	18.7	1151	9	US-09-350-259-37	Sequence 37, Appl
40	1108	18.7	1151	10	US-09-891-943-37	Sequence 37, Appl
41	1091	18.4	1153	9	US-09-350-259-3	Sequence 3, Appl
42	1091	18.4	1153	10	US-09-902-481A-1	Sequence 1, Appl
43	1091	18.4	1153	10	US-09-891-943-3	Sequence 3, Appl
44	1091	18.4	1153	14	US-10-144-259-30	Sequence 30, Appl
45	1091	18.4	1153	14	US-10-207-655-176	Sequence 176, App

ALIGNMENTS

RESULT 1

US-10-741-601-531
; Sequence 531, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01500
; CURRENT APPLICATION NUMBER: US/10/741.601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 531
; LENGTH: 1167
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-531

Query Match 99.3%; Score 5890.5; DB 16; Length 1167;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1129; Conservative 2; Mismatches 1; Indels 35; Gaps 1;

Qy	1	MELPFVTHLFLPLVLTGLCSPNLDHHPRLPGPPEAEFGYSLQHVGGQRMVLGA	60
Db	1	MELPFVTHLFLPLVLTGLCSPNLDHHPRLPGPPEAEFGYSLQHVGGQRMVLGA	60
Qy	61	PWDGSDRRGDVYRCVPVGAHNAKAGHLDYQLGNSHPAVNMHLGMSLLTDDGGG	120
Db	61	PWDGSDRRGDVYRCVPVGAHNAKAGHLDYQLGNSHPAVNMHLGMSLLTDDGGG	120
Qy	121	FMACAPLWSRACSSVFFSGICARVDASFPQGSAPTAQRCPPTMDVIVLDGNSNIYP	180
Db	121	FMACAPLWSRACSSVFFSGICARVDASFPQGSAPTAQRCPPTMDVIVLDGNSNIYP	180
Qy	181	WSEVQTFRLRLVKGKLFIDPEQIQVGLVQYGESPVHWSLGLDFRTKEEVRAAKNLSREG	240

```
Db 181 WSEVQTLRLRLVGLFDPQIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSREG 240
Qy 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVGYDMGGSVLMLEGGHRLPPRMALEDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVGYDMGGSVLMLEGGHRLPPRMALEDEF 420
Qy 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPFRHRGKVIAPQLKXDGAVRVAQSLOGE 480
Db 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPFRHRGKVIAPQLKXDGAVRVAQSLOGE 480
Qy 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTLLQ 540
Db 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTLLQ 540
Qy 541 PEPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGALYLYHGTSQGVRRPHPAORIA 600
Db 541 PEPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGALYLYHGTSQGVRRPHPAORIA 600
Qy 601 AASMPHALSYFGRSVDRGLDLDGDDLVDVAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSYFGRSVDRGLDLDGDDLVDVAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660
Qy 661 VVQDCCRROGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Db 661 VVQDCCRROGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSRRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLINEGSPSIIQ 780
Db 721 RLSRRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLINEGSPSIIQ 780
Qy 781 KLVPFKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKVLVSTTLNKRKNAYNT 840
Db 781 KLVPFKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKVLVSTTLNKRKNAYNT 840
Qy 841 SLSIFSRNLHLASLTTPQRESPIKVECAAPSAHARLCSVGHVPVQTCAGKVTFLLEFEPSC 900
Db 841 SLSIFSRNLHLASLTTPQRESPIKVECAAPSAHARLCSVGHVPVQTCAGKVTFLLEFEPSC 900
Qy 901 SLLSQQVFGKLTASSDSLERNGTLQENTAOATSAYIQIEPHLLPSSSESTLHRYEVPYGTLL 960
Db 901 SLLSQQVFGKLTASSDSLERNGTLQENTAOATSAYIQIEPHLLPSSSESTLHRYEVPYGTLL 960
Qy 961 PVGPGPSEKFTTLR-----TNNASCIVQNLT 985
Db 961 PVGPGPSEKFTTLRVQNLGCVYVSGLIISALLPAVAHGNVFLSQQVITNNASCIVQNLT 1020
Qy 986 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLQAKGTESVGLLRLVHNEFFRRAKFK 1045
Db 1021 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLQAKGTESVGLLRLVHNEFFRRAKFK 1080
Qy 1046 SLTVSTVFELGTREGSVLQJTEASRWSESLEVVQTRPILISWILGSLVGLLLALL 1105
Db 1081 SLTVSTVFELGTREGSVLQJTEASRWSESLEVVQTRPILISWILGSLVGLLLALL 1140
Qy 1106 VFCLWKLGFPAHKKIPBEEKREKLEQ 1132
Db 1141 VFCLWKLGFPAHKKIPBEEKREKLEQ 1167
```

RESULT 2

US-10-741-601-532

; Sequence 532, Application US/10741601

; Publication No. US20040166519A1

; GENERAL INFORMATION:

```
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/10/741,601
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-532
```

Query Match 99.3%; Score 5890.5; DB 16; Length 1177;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1129; Conservative 2; Mismatches 1; Indels 35; Gaps 1;

```
Qy 1 MELPFFVTHLPLVFLVLTGLCSPFNLDHHPRLPPGPEAFYGVSLQHVGGQRMVLVGA 60
Db 1 MELPFFVTHLPLVFLVLTGLCSPFNLDHHPRLPPGPEAFYGVSLQHVGGQRMVLVGA 60
Qy 61 PWDGPGSDRRGDIYRCPVGAHNAAPCAKGLHLDYQLGNSHHPAVNMHLGMSLLETDDGG 120
Db 61 PWDGPGSDRRGDIYRCPVGAHNAAPCAKGLHLDYQLGNSHHPAVNMHLGMSLLETDDGG 120
Qy 121 FMACAPLMSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPTYMDVTVLVDGNSIYP 180
Db 121 FMACAPLMSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPTYMDVTVLVDGNSIYP 180
Qy 181 WSEVQTLRLRLVGLFDPQIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSREG 240
Db 181 WSEVQTLRLRLVGLFDPQIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSREG 240
Qy 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVGYDMGGSVLMLEGGHRLPPRMALEDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVGYDMGGSVLMLEGGHRLPPRMALEDEF 420
Qy 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPFRHRGKVIAPQLKXDGAVRVAQSLOGE 480
Db 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPFRHRGKVIAPQLKXDGAVRVAQSLOGE 480
Qy 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTLLQ 540
Db 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTLLQ 540
Qy 541 PEPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGALYLYHGTSQGVRRPHPAORIA 600
Db 541 PEPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGALYLYHGTSQGVRRPHPAORIA 600
Qy 601 AASMPHALSYFGRSVDRGLDLDGDDLVDVAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSYFGRSVDRGLDLDGDDLVDVAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660
Qy 661 VVQDCCRROGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Db 661 VVQDCCRROGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSRRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLINEGSPSIIQ 780
Db 721 RLSRRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLINEGSPSIIQ 780
Qy 781 KLVPFKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKVLVSTTLNKRKNAYNT 840
Db 781 KLVPFKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKVLVSTTLNKRKNAYNT 840
```

```
Qy 841 SLSIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVQTGAKVTFLLEPFSC 900
Db 841 SLSIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVQTGAKVTFLLEPFSC 900
Qy 901 SLSLSQVFGKLTASSDSLRNGTLOENTAOISAVIQYEPHLLFSSSTLHRYEHPYGTLL 960
Db 901 SLSLSQVFGKLTASSDSLRNGTLOENTAOISAVIQYEPHLLFSSSTLHRYEHPYGTLL 960
Qy 961 PVGPGPFKTLR-----TNNASCIVQNL 985
Db 961 PVGPGPFKTLRQNLGCVYVSGLIISALLPAVAHGGNYFLSLSQVITNNASCIVQNL 1020
Qy 986 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLQOLAKGTEVSGILRLVHNEPFRRAKFK 1045
Db 1021 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLQOLAKGTEVSGILRLVHNEPFRRAKFK 1080
Qy 1046 SLTVVSTFELGTEGSLVQLTEASRWSESLLEVQTRPILISLWILGSLGLLALL 1105
Db 1081 SLTVVSTFELGTEGSLVQLTEASRWSESLLEVQTRPILISLWILGSLGLLALL 1140
Qy 1106 VFCLWKLGFPAHKKIPBEKEEKLQ 1132
Db 1141 VFCLWKLGFPAHKKIPBEKEEKLQ 1167

RESULT 3
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338

Query Match 39.7%; Score 2355; DB 15; Length 1188;
Best Local Similarity 42.3%; Pred. No. 2.6e-199;
Matches 506; Conservative 204; Mismatches 406; Indels 80; Gaps 16;

Qy 1 MELPFVTHLFLVFLTLGLCSPPFNLDHHPFLPGPPEAFEGSYVLQHVGGQRMVLVGA 60
Db 1 MDLPRGLVVAWALSILWPGFTDTFNWDTKPRVIFGSRTPAFGYTVQHQHDSGNKWL VGA 60
Qy 61 PWDGPGDRRGDVRCPVGGAAHAPCAKHGLGDYQLGNSSHPAVNMHIGMSLLETGDGG 120
Db 61 PLETNGYQKTGDVYKCPV---IHGNTCKNLGRVTLSNVSRKDNMRGLSLATNPXDS 117
Qy 121 FMACAPLWSRAGSSVFSSGICARVDASFQGGSLAPTAQRCPTVMDVIVLDGNSIYP 180
Db 118 FLACSPLWSHCEGSSYTTGCSRVNSNFRSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Qy 181 WSEVQTFRLRLVGLKFLDIPDQIQVLGVQYGSPPHVEWSLGDGFRTKEEVVVRAAKNLSREG 240
Db 178 WVEVQHFLINILKKFYIGPGQIQGVVQYGEDVHVEHFLNDYRSVKDVEASHIEQRGG 237
```

```
Qy 241 RETKTAQIMVACTEGFSQSHGGRPEARLLVVVTGDSHGDGELPAALKACEAGRVTY 300
Db 238 TETRTAFGIEFAFSEAFQK--GGRKGAKKVMIVITDGSHDSPLEKVIQOESRDNVTY 295
Qy 301 GIAVLGHYLRQRDPDFSLREIRTIASDPDERFFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYNNRGINPETFLEIKYIASDDPKDHFNFVNTDEAALKDIVDALGDRIFSLEGT 355
Qy 361 HANNESSFGLMSOIGFSTHRLKDGILFGMVGYADWCGSVLWLEGHRLPPRMALEDEF 420
Db 356 N-KNETSFGLMSQTGFSSHVVEDGVLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414
Qy 421 PPALONHAAVLYGSVSMLLRGGRRLLFLSGAPRFRHKGKVIAPOLKKGAVRVQAQSLQGE 480
Db 415 PEELKNHAGLYGTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSITUIQAMRGQ 474
Qy 481 QIGSYFGSELCPDTRDGTDLVLLVAAPMFLGPNQKQKTRVYVYLVGQQSLLTQGTLLQ 540
Db 475 QIGSYFGSEITSVDIDGQDVTDLVVGAPMYFN--EGRGKGVYVYEL--RQNFVYNGTLK 532
Qy 541 PEPP-QDARFGFANGALPDLNODGFADVAVGAPLEDHOGALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVWVVGAPLEDNHAGAIYIFHGRGSLKTPKQRI 592
Qy 600 AAASMPHALSYFGSSVDGRLLDGDGLVDVAVAGGGAAILLSRPVHLTPSLEVTPOAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNAILWSRPVVOINASHLFEPSKI 652
Qy 660 SVQORDCRRRGQEAVALCTAALCFQVTSRTPGMDHQFYMFRFTASLDEWATAGARAADGSG 719
Db 653 NIHRDCKRGRDATCLAAFLCTPIFLAHFQITTVGIRYNATMDERRYTTPRAHLDBG 712
Qy 720 QRLSPRRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTVFALDNTTKPGVNLGSPFSI 779
Db 713 DRFTNRAVLSSGQELCERINFHVLDTADYVVKPVTFSEVSESLDPDH--GPMLDGWPPTL 771
Qy 780 QKLVPFKDCGPDNECVTDVLQVNMDI-----RGRK-----APVW 817
Db 772 RVSVFPMWNGCNEDEHCVDPDLVDARSDLPTAMEYCORVLRKPAQDCSAYTSLFDTTVEII 831
Qy 818 RGRKRVLVSTTLNRKENAYNTSLSIIFSRNLHLASLTPORESPIKVECAAPS--AHAR 875
Db 832 ESTFORVAVEATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSGIECVNEERLQKQ 891
Qy 876 LCSVGHVPVOTGAKVTFLLEPFSCSLLSQVFKLTASSDSLRNGTLOENTAOISAYI 935
Db 892 VCNVSYPPFRAKAKAVAFRLDPEFSKIFLHLHLELAAGSDSNERDSTKEDNAPLRFHL 951
Qy 936 QYEPHLLFSSSTLHRYEHPYGTLP--VGPGPEFKTLRNL----- 975
Db 952 KYEADVLFTRSSLSHYEVKPNSSSLRYDYGIGPPFSCIFRIONLGLPFIHGMKMITIPI 1011
Qy 976 -----NASC-IVQNLTPEPPGPPVHPELOHTNRLNGSNTCCQVVR 1015
Db 1012 ATRSGNRLKLRLDPLTDEANTSCNIWGNSTYRTPVE--EDLRAPQLNHSNDVVSINC 1070
Qy 1016 HLQOLAKGTEVSGLLRLVHNEPFRRAKFSLTAVVSTFELGTEGSLVQLTEASRWSSSL 1075
Db 1071 NI-RLVNEQIEINPHLLGNLWLRSLKALKYKSMKIMVNAALQORQPHSPPIFREDDPSRQIV 1129
Qy 1076 LEVVQTRPILISLWILGSLGLLALLVFLWKLGFPAHKKIPBEKEEKLQ 1131
Db 1130 FEISKQEDWQVPIIIVGSLTIGLLALLVALLWKLGFPRASR-----RRREPGLD 1181

RESULT 4
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
```

[illegible]

```
Db 118 FLACPLWSHECGSSYTTGMCNRVNSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Qy 181 WSEVQTFLRLRLVGLKFLDIPQIOVLGVQYSGSPVHWSLGDFTKKEEVRAAKNLSREG 240
Db 178 WVEVQHFLINILKFKYIGPGQIQGVVQYGEDVVHFFHLNDYRSVKDVEAASHIEQRGG 237
Qy 241 RETKTAQIIVACTEGFSQSHGGRPEARLLVVVTDGSHDGEELPAALKACEAGRTRY 300
Db 238 TETRTAFGIFARSEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOOSRDNVTRY 295
Qy 301 GIAVLGHLRRQDRPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYYNRGINPETFLNEIKYIASDDPKHFFNVTDAAALTDIVDALGDRIFLSLEG 355
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQTFGSHVVEDGVLGAVGAYDMNGAVLKETSACKVPLRESYLKEF 414
Qy 421 PPALQNHAAVLGYSVSMLLRGRRFLSGAPRFRHKGKVIATOLKKDGAVRVAOSLOGE 480
Db 415 PEBLKNHGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSLLTIHQAMRQ 474
Qy 481 QIGSYFGSELSCLPTDRDGTDLVLLVAAPMFLGPQNKETGRVYVVLVQOQSLLTLQGTLO 540
Db 475 QIGSYFGSEITSVDIDGQDVTDLVLCAPMYFN--EGREKGVYVEL--RQNRFYVNGTLK 532
Qy 541 PEPP-QDARFGFAMGALPDNLNQGFADVAVGAPLEDGHQAGLYLHYGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPKQRI 592
Qy 600 AAASMPHALSYFGSRVDGRDLDDGLVDVAVAGAGAAIILSSRPVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSRPVQINASHLFEPSKI 652
Qy 660 SVVQDCCRREGOEAVCLTAALCFQVTSRTPCRWHDQFYMFRTASLDEWTAGARAFDGS 719
Db 653 NIFHRDCKSRGRDATCLAAFLCTPIFLAPHFQTTVGIIRYNATMDERRYPTRAHLDEGG 712

Db 118 FLACPLWSHECGSSYTTGMCNRVNSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Qy 181 WSEVQTFLRLRLVGLKFLDIPQIOVLGVQYSGSPVHWSLGDFTKKEEVRAAKNLSREG 240
Db 178 WVEVQHFLINILKFKYIGPGQIQGVVQYGEDVVHFFHLNDYRSVKDVEAASHIEQRGG 237
Qy 241 RETKTAQIIVACTEGFSQSHGGRPEARLLVVVTDGSHDGEELPAALKACEAGRTRY 300
Db 238 TETRTAFGIFARSEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOOSRDNVTRY 295
Qy 301 GIAVLGHLRRQDRPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYYNRGINPETFLNEIKYIASDDPKHFFNVTDAAALTDIVDALGDRIFLSLEG 355
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQTFGSHVVEDGVLGAVGAYDMNGAVLKETSACKVPLRESYLKEF 414
Qy 421 PPALQNHAAVLGYSVSMLLRGRRFLSGAPRFRHKGKVIATOLKKDGAVRVAOSLOGE 480
Db 415 PEBLKNHGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSLLTIHQAMRQ 474
Qy 481 QIGSYFGSELSCLPTDRDGTDLVLLVAAPMFLGPQNKETGRVYVVLVQOQSLLTLQGTLO 540
Db 475 QIGSYFGSEITSVDIDGQDVTDLVLCAPMYFN--EGREKGVYVEL--RQNRFYVNGTLK 532
Qy 541 PEPP-QDARFGFAMGALPDNLNQGFADVAVGAPLEDGHQAGLYLHYGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPKQRI 592
Qy 600 AAASMPHALSYFGSRVDGRDLDDGLVDVAVAGAGAAIILSSRPVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSRPVQINASHLFEPSKI 652
Qy 660 SVVQDCCRREGOEAVCLTAALCFQVTSRTPCRWHDQFYMFRTASLDEWTAGARAFDGS 719
Db 653 NIFHRDCKSRGRDATCLAAFLCTPIFLAPHFQTTVGIIRYNATMDERRYPTRAHLDEGG 712
```

RESULT 6

US-09-836-353A-35

```
; Sequence 35, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PP489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-836-353A-35
```

```
Query Match 39.5%; Score 2344.5; DB 10; Length 1189;
Best Local Similarity 42.2%; Pred. No. 2.3e-198;
Matches 505; Conservative 204; Mismatches 407; Indels 81; Gaps 16;

Qy 1 MELPFTVTHLPLVFLVLTGLCSPPNLDDEHPRLPGPPEAFYGSVLQHVQGGQRMMLVGA 60
Db 1 MDLPRLGVAAWALSMPGFTDTFNMDTRKPRVPGSRTAFPGYTVQQHDSIGNKWLVVGA 60
Qy 61 PWDPGSDRRGDVYRCPVGGAHNAPCAKGLHLDYQLGNSHHPAVNMHLGMSLLETDDGG 120
Db 61 PLETNGYQKTDYVYKCPV---IHGNCCTKLNLGRVLTNSVSEKDNMRLGLSLATNPKDNS 117
Qy 121 FMACAPLWSRACSSVSSGICARVDASFOQGS LAPTAQRCPTYMDVIVLDGNSNIYP 180
Db 118 FLACPLWSHECGSSYTTGMCNRVNSFRFSKTVAPALQRCQTYMDIVIVLDGNSNIYP 177
Qy 181 WSEVQTFLRLRLVGLKFLDIPQIOVLGVQYSGSPVHWSLGDFTKKEEVRAAKNLSREG 240
Db 178 WVEVQHFLINILKFKYIGPGQIQGVVQYGEDVVHFFHLNDYRSVKDVEAASHIEQRGG 237
Qy 241 RETKTAQIIVACTEGFSQSHGGRPEARLLVVVTDGSHDGEELPAALKACEAGRTRY 300
Db 238 TETRTAFGIFARSEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOOSRDNVTRY 295
Qy 301 GIAVLGHLRRQDRPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYYNRGINPETFLNEIKYIASDDPKHFFNVTDAAALTDIVDALGDRIFLSLEG 355
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQTFGSHVVEDGVLGAVGAYDMNGAVLKETSACKVPLRESYLKEF 414
Qy 421 PPALQNHAAVLGYSVSMLLRGRRFLSGAPRFRHKGKVIATOLKKDGAVRVAOSLOGE 480
Db 415 PEBLKNHGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSLLTIHQAMRQ 474
Qy 481 QIGSYFGSELSCLPTDRDGTDLVLLVAAPMFLGPQNKETGRVYVVLVQOQSLLTLQGTLO 540
Db 475 QIGSYFGSEITSVDIDGQDVTDLVLCAPMYFN--EGREKGVYVEL--RQNRFYVNGTLK 532
Qy 541 PEPP-QDARFGFAMGALPDNLNQGFADVAVGAPLEDGHQAGLYLHYGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPKQRI 592
Qy 600 AAASMPHALSYFGSRVDGRDLDDGLVDVAVAGAGAAIILSSRPVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSRPVQINASHLFEPSKI 652
Qy 660 SVVQDCCRREGOEAVCLTAALCFQVTSRTPCRWHDQFYMFRTASLDEWTAGARAFDGS 719
Db 653 NIFHRDCKSRGRDATCLAAFLCTPIFLAPHFQTTVGIIRYNATMDERRYPTRAHLDEGG 712
```

```
Qy 720 QRLSPRLRLSVGNVTCEQLHFHVLDSYLRPVVALTVTFALDNTTKPGFVLNEGSPTSI 779
Db 713 DRFTNRAVLSSGQELCERINFHVLDTADYVKVPTFSVEYSLEDPDH-GPMLDDGWPTTL 771
Qy 780 QKLVPPSKDCPDNECVTDVLQVNMDI-----RGRK-----APFVV 817
Db 772 RVSVPFWNGCNEDEHCVPLDLVLAARSDLPTAMEYQCVLRKPAQDCSAYTLSEFTTVFII 831
Qy 818 RGRKRVLVSTLENKENAYNTSLIFSRNLHLASLTQORSPPIKVECAAPS--AHAR 875
Db 832 ESTRQRVAVATLENRENAYSTVLNISQSANQFASLIQKEDSDGSEICNEERRLOKQ 891
Qy 876 LCSVGHVPVFTGAKVTELLBFEPFSCSLLSQVFGKLTASSDSLSRNGTLOENTTAQTSAYI 935
Db 892 VCNVSYPFRAKAKVAFRLDFEFSKSIPLHLELELAAGSDSNERDSTKEDNVAFLRPHL 951
Qy 936 QYBPHLLFSSESTLHRYEVHPYGTLP--VPGPBPKTTLRTN----- 975
Db 952 KYEADVLFTRSSLSHYEVKLNLSLERYDGTGPPFCIFRIQNLGLPFIHGMKTIPI 1011
Qy 976 -----NASC-IVONLTPPGPPVHPBELOHTNRLNGSNTQCQVR 1014
Db 1012 ATRSGNLLKLDRFLTDEVANTSCNIWNGSTERYPTVE-EDURRAPOLAHNSDVVVSIN 1070
Qy 1015 CHLQOLAKGTEVSVGLLRLVHNEFFRRAKPKSLTVSTFELGTBEGSVLQ/LTEASRWSES 1074
Db 1071 CNI-RLVPNQEPHLLGNLWLSLKALKYKSMKIMYNAAALQRFHSPFIFREEDPSRQI 1129
Qy 1075 LLEVQTRPILISWILIGSVLGGLLLALLVFCWLKLGFFAHKXIPEEKREBKELE 1131
Db 1130 VFEISKQEDQVPIWIIVGTSLGGLLLALLVLALWKLGFPSAR----RRREPGLD 1182

RESULT 7
US-10-262-839-4
; Sequence 4, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsebrook, David W.,
; APPLICANT: Anderson, John,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Kekuda, Rameah,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shinkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR FILING DATE: 2002-10-01
; PRIOR FILING DATE: 2001-10-02
; PRIOR FILING DATE: 2001-10-09
```

```
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 4
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-262-839-4

Query Match 39.5%; Score 2344.5; DB 15; Length 1189;
Best Local Similarity 42.3%; Pred. No. 2.3e-198;
Matches 505; Conservative 204; Mismatches 407; Indels 81; Gaps 16;

Qy 1 MELPFTVTHLPLVLFTGLTGLCSPPNLDHHPRLPPGPEABFGYSVLQVHGVGGGRWMLVGA 60
Db 1 MDLPRLGVVAWALSMLPFGFTDTFNMMDTRKPRVIPGSRTPAFGYTVQOHDISGNKWLTVGA 60
Qy 61 PWDGPGDREGDYVRCVPGCAHNAKPGKHLGDIYQVQVGNSSHPVNMHLGMSLLETDCDGG 120
Db 61 PLETNQYQKTDGDIYKCPV---IHGNCVKLMLGRVTLNSVSRKDNMRGLGSLATNPKNDS 117
Qy 121 FMACAPLWSRACSSVFSFGICARVDASFOQOGLAPTAQRCPTQMDVTVLDGNSIYP 180
Db 118 FLACSLMSHECGSSVYTCMCSSRNSNFRSKTVAPALQRCQTYMDIVLDGNSIYP 177
Qy 181 WSEVQTPRLRLVKLFIDPEQIOVLGVQYESPVHWSLGDFTKEBVEVVAANKLSRREG 240
Db 178 WVEVQHFLINILKKFYIGPQIQVGVVQYGEDVHFEHLNDYRSVDVVEAASHIEQRG 237
Qy 241 RETKTAQIMVACTEGFSQSHGGRPEARLLVVVTGESHGDEELPAALKACEAGRTRY 300
Db 238 TETRTAFGIEFARSEAFQK--GGRKGAKVMIVITGESHSDSPDLEKVIQOOSERDNTRY 295
Qy 301 GIAVLGHYLRQRDPSPFLREIRTIASDDPDERPFFNVTVDEAALTDIVDALGDRIFGLGEGS 360
Db 296 AVAVLGYNNRRGINPETFLNEIKYIASDDPKHFFNVTVDEAALKDIVDALGDRIFSLGEGT 355
Qy 361 HARNESFGLMSQIGFSTHRLKDGILFGMVAYDMGSGVLMLEGHRLFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQTFSSHVEDGVLGAVGAYDMGAVLKVETSAGKVIPLRESYLKEF 414
Qy 421 PPALQNHAAVLGYSVSSMLLRGGRRLLSLGAPFRHGRKVIAFOLKDGAVRVVAQSLQGE 480
Db 415 PEELKNHGYLVGYTVTSVSSROGRVYVAGAPRPNHTGKVLFTFMHNNRSLTIHQAMRGQ 474
Qy 481 QIGSYFGSELCPDTRDGTDTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOQSLTLQGTTLQ 540
Db 475 QIGSYFGSEITSDVIDDGDGVTDLVVGAPMYPN-EGREKGVVYVEL-RQNRVYVNGTLK 532
Qy 541 PEPP-QDARTGFAMGALPDNLNQGFADVAGAPLEDHQGALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQSDYNDVVVVGAPLEDNHAGAIYIFHGFSGSILKTPKQRI 592
Qy 600 AASMPHALSYRGSVDGLDLDGDDLDVAVAGQAAILLSRPIVHLTPSLVLTPOAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLDLDVAVAGNAVILMSRPVVQINASLHFEPSKI 652
```



```

; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: P489P1
; CURRENT APPLICATION NUMBER: US/09/836.353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-43

Query Match 37.0%; Score 2194; DB 10; Length 1034;
Best Local Similarity 44.3%; Pred. No. 4.3e-185;
Matches 460; Conservative 178; Mismatches 346; Indels 54; Gaps 13;

Qy 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPGLPPEABFGYSLVQLHVGQGRWMLVGA 60
Db 1 MDLPRGLVAVWALSLWPGFTDTFNMNDRKPRVIFGSRTPAFPGYTVQQHDISGNKWLWVGA 60
Qy 61 PWDPGSDRGDVRCPVGGNHNAPCAKGLHGLDQLGNSHPAVNMHLGMSLLETDDGG 120
Db 61 PLETFNGYQKTDVYKCV--IHNCCKNLNIGRVTLSNVSERKDNMRLGLSLATNPKNDS 117
Qy 121 FMACAPLWSRACGSSVSSGICARVDASFOQSGSLAPTAQRCPYMDVIVLDSGSIYP 180
Db 118 FLACPLWSHCEGSSYTTGCSRNSNFRPSKTVAPALQRCQTYMDIVIVLDSGSIYP 177
Qy 181 WSEVQTFRLRLVGLKFLIDPEQIQVLQVYGSPPVHWSLGDPRTKBEEVRAAKNLSRREG 240
Db 178 WVEVQHEFLINLKFKYIGPGQIQGVYQGVDDVHFEHLNDYRSVKDVEAASHIEQRG 237
Qy 241 RETTAQAINVACTEGSQSHGGRPEARLLVVVTDGESHGDELPALAKACEAGRTRY 300
Db 238 TETATFGEIPEARSEAFQK--GGRKGAKVMIVITDGHSDSPDLEKVIQOOSRDNVTRY 295
Qy 301 GIAVLGHLRRORDPSFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYNRGINPEFLNEIKVIASDPDKHFNVTDEAALXDIVDALGDRIFSLGEGT 355
Qy 361 HAENESSFGLMSQIGFSTRHLKDKGILFGVMGAYDWGWSVLWLEGHRLPFPRLMALEDEF 420
Db 356 N-KNETSFGLMSQTFSSHVVEDGVLGAVGYDMNGAVLKETSAGKVIPLRESYLKEF 414
Qy 421 PPALONHAAYLGYSVSSMLLGGRRFLSGAPRRHRGKVIAFOLKDGAVRVAQSLQGE 480
Db 415 PEELKNHAGLYGTVTVSVSSRQGRVYVAGAPRNHTGKVLFTMHNRSLSITHQAMRGQ 474
Qy 481 QIGSYFGSELCPDLDRDGTDLVLLVAMPFLGPQNKETGRVYVYLVQVQOGLLTQGLTQ 540
Db 475 QIGSYFGSEITSDIDGGVTDVLLVAGPMYFN--EGERGKYYVEL--RQNRFYVNGTLK 532
Qy 541 PEPP-QDARFGFANGALPDNLQDGFADVAGAPLEDHQGALYLYHGTSQGRVPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNODSYNDVVVGAPELNDHAGAIYIFHGFPGSILKTPQRI 592
Qy 600 AASMPHALSVFGRSDGRDLDDGDLVDVAVGAQCAAILLSRPIVHLTPSLVETQAI 659
Db 593 TASELATGLQYFGCSIHGQLDNEGLDIDLVAGLGNVILWSPVQVQINASLHFEPSKI 652
Qy 660 SVQORDRRRQGEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAAPDQSG 719
Db 653 NIFHRDCKSRDATCLAAFLCPTPIFLAPFQTTTGVIRNATMDEKXYTPRAHLEGG 712
Qy 720 QRLSPRLRLSVGNVTCQLHFLHVLDTSDYLRPVALTVTFALDNTTKPGPVLNEGSPTSI 779
```

```

Db 713 DRFTNRAVLLSSQOELCERINFHVLDTADYVKVETFSVEYSLSDPDH-GPMLDDGWPTTL 771
Qy 780 QKLVPPSKCGPDNECVTDLVLQVNMDI-----RGSRK-----APFVV 817
Db 772 RSVVPFWNGCNEDEHCVDPDLVLDARSDLPTAMBYCORVLKPKAQDCSAYTLSPDTTVFII 831
Qy 818 RGGRRKVLVSTTLERKENAYNTSLSIIFSRNLHLASLTPQRESPIKVECAAPS--AHAR 875
Db 832 ESTRQRVAVENTLENGENAYSTVLNISQSANLQFASLIQKEDSGSIECVNEERLQKQ 891
Qy 876 LCSVGHVPFOTGAKVTFLLEFPSCSLLSQVFGKLTASSDSLSRNGTLQENTAQTSAYI 935
Db 892 VCNVSYPPFFRAKAKVAFRLDFFESKIFLHLHLEILAAGSDSNERSDSTKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSESTLHRYEVHPYCTLP--VCPGPEPKTTLRTNNASCIVQLNTSPPGPPVH 993
Db 952 KYEADVLFTRSSLSLHYEVKLNSSLSRYDGIQGFPPFICFRIQNLGLP-----PIH 1001
Qy 994 -----PEELQHTNRL 1003
Db 1002 GIMMKITIPATRSNRL 1019

RESULT 10
US-10-262-839-6
; Sequence 6, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voess, Edward,
; APPLICANT: Zernhosen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-452A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
```



```
Qy 255 EGFSSQSHGPEAARLLVVTVDGESHDBELPAALKAACEAGRVTRYGIADVGLHYLRQRD 314
Db 260 EAFTAGARRGVKVMVITVDGESHDBELPAALKAACEAGRVTRYGIADVGLHYLRQRD 319
Qy 315 PSSFLRIRTIASDPDRFPFNVTDEAALTDIVDALGDRIFGLGSHAEENESSFGLMSQ 374
Db 320 TEKPFVEIKSIASPTKHFNFVSDDELAVTIYKTLGERIFALEATADQSAASFEMMSQ 379
Qy 375 IGFSTRHLKDGILFGMGVADWGSVLMWEGGHRLLPPRMALEDEFPALQNAHYLGYS 434
Db 380 TGFSAHYSQDWMVLGAVGADWGTVMVQKASQIIIPNTTFNVSTKNEPLASYLGYT 439
Qy 435 VSSMLLGGRRLLFLSGAPRFRHRGKVIATFQKDGAVRAQSLOGEIGSYFSGELCPD 494
Db 440 VNSATSSGDVLYTAGQPRYNTQVVIYRM-EDGNIKILQTLGSGEIGSYFSGILTTD 498
Qy 495 TDRGTTDVLVAAAPMFLGPNKQKTRGVYVYLVGQQSLLTLQTLQP----- 541
Db 499 IDKDSNTDILLVGAPMYMGTEKEGKGVYVYAL-NOTRPFYQMSLEPIKOTCCSSRQHS 557
Qy 542 -----BPPQDARFGFANGALPDNLQDGFADVAGAPLEDHOGALYLYHGTQSGVRPH 595
Db 558 CTTENKNEPCGARTAIKAVKDLNLDGFNDIVIGAPLEDHOGAVYIYHSGKTIKEY 617
Qy 596 AQRIAAASMPHALSYFGORSVDRLDGLDGLDVAVAGAAQAAILLSSRPVHLTPSLEV 655
Db 618 AQRIPSGGDKTKLKFQSGIHGMDLNGDGLTDVTIGLGAALFWSRDVAVVKVWTFE 677
Qy 656 PQAISVVQDRCRRGQAVCLTAALCFQVTSRTPGRWDPHQPMTASLDEWTAGARAF 715
Db 678 PNKVNIOKKCHMEGKETVCINATVCPKSKEDTIYEADLQRYVTLDSLRSRFF 737
Qy 716 DGSQRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTVPALDNTTKP--GPVLNE 773
Db 738 SGTQERKQVR--NITVRKSECTKHSFYMLDKHDFQDSVRITLDP---NLTDPEMGPVLD 792
Qy 774 GSPTSIOKLVPFSKDCPDNECVTDVLQVNMIDIRSKAPFVVRGGRKRVLVSTTLNR 833
Db 793 SLPNVHEIYIPFAKDCGKKEKICISLSLV---ATTEKDLIIVRSQNDKFNVSITVKNT 848
Qy 834 KENAYNTSLIIPRNHLASLTQRESPIKVECAAPSAHARLCSVGHVPQCAKUTFL 893
Db 849 KDSAYNTRTIYHSPNLVFSGI-----EAIQKDSK--ESNHNITCKVGPFLRRGEMVTFK 902
Qy 894 LEPFSCSSLLSQVFGKLTASSDLSRNGTLQENTAOISAVIOYEPHLLPSESSTLHRYE 953
Db 903 ILFOFNTSYLMENVTIYLSATSDSEPPETLSDNVNVNISIPKVEVLQFVSSASEHIS 962
Qy 954 VHPYGTLP-----VG-----PCPEFKTTLRTNNASCIVQNLTPEPPGP 990
Db 963 IAAENETVPEVINSTEDIGNEINIFYLRKSGSPMPELKLIS-----PPNMTSNGYP 1015
Qy 991 PVHPEELQHTNRLN-----GNTQCO-----VVRCHLG 1018
Db 1016 VLYPTGLSSSENACRPHIPEDPFSSINSKGKMTTSTDHLKRGTTLDNCNCKFATITCNLT 1075
Qy 1019 QLANGTEVSGLLRLVNERPRRAKFKSLTVVSFFELGTBEGSVLQLTAEASRESSELEV 1078
Db 1076 S-SDISQVNSL---ILWKPTFKSYFSSNLTIIGELRSENAS-LVLSSSKQKELAIQI 1131
Qy 1079 VQT-RPILISLWILIGVLGGLLALLAVFLKWLGFPAHKKIPBEKREE 1128
Db 1132 SKDGLPGRVPLVILLSAFAGLLMLLILALWKIGFF---KRLPKKQMEK 1179
```

RESULT 12

```
US-09-918-715-307
; Sequence 307, Application US/0918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
```

```
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 307
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Rat
; US-09-918-715-307

Query Match 31.8%; Score 1884; DB 10; Length 1180;
Best Local Similarity 36.2%; Pred. No. 2e-157;
Matches 432; Conservative 219; Mismatches 432; Indels 112; Gaps 25;

Qy 13 LVFLTGLCSPPNLDEHHPLFPQPPEASFGYSVLQHVGGQRMVLGAPWDGSGDRGD 72
Db 19 LVFLTGLCSPPNLDEHHPLFPQPPEASFGYSVLQHVGGQRMVLGAPWDGSGDRGD 78
Qy 73 VYRCPVGAHNAFCAGHGLGDYQLGNSHPAV----NMHLGMSLLETDCDGGCFMACPL 127
Db 79 VYRCPVGAHNAFCAGHGLGDYQLGNSHPAV----NMHLGMSLLETDCDGGCFMACPL 133
Qy 128 WSRAGSSVSSGICARVDASFPQGLSAPTAQRCPTYMDVIVLDGNSIYPMSEVQTF 187
Db 134 YAYRCHLHVTTCGSDVSPTQVNSFAP-VQECSTQLDIVLDGNSIYPMSEVQTF 192
Qy 188 LRLVGLKFLDPIQIQLVQGESPHWESLGDFTKEBVRVAAKLSRREGRETAK 247
Db 193 LNDLLKRMIDIGPKQTQVIGVGENVTHEFNLNKYSSTEEVVAANKIGRGGQLTMTAL 252
Qy 248 AIMVACTGFGSQSHGGRPEAARLLVVVTVDGESHDBELPAALKAACEAGRVTRYGI 307
Db 253 GIDTARKEATFAGARGGVKVMVITVDGESHDBELPAALKAACEAGRVTRYGI 312
Qy 308 YLRQRDPSSFLRIRTIASDPDRFPFNVTDEAALTDIVDALGDRIFGLGSHAEENESS 367
Db 313 YNRGNLSTKFEVSEIKSIASEPTKHFNFVSDDELAVTIYKTLGERIFALEATADQSAAS 372
Qy 368 FGLMSQIGSTRHLKDGILFGMGVADWGSVLMWEGGHRLLPPRMALEDEFPALQNH 427
Db 373 FEMMSQTFSAHYSQDWMVLGAVGADWGTVMVQKASQIIIPNTTFNVSTKNEPLASY 430
Qy 428 --AAYLGYSVSSMLLGGRRLLFLSGAPRFRHRGKVIATFQKDGAVRAQSLOGEIGSY 485
Db 431 PLASYLGYTVNSATIPGD-VLYTAGQPRYNTQVVIYRM-EDGNINILQTLGGEQIGSY 488
Qy 486 FGSSELCPDTRDGTDDVLVAAAPMFLGPNKQKTRGVYVYLVGQQSLLTLQTLQP---- 541
Db 489 FGSVLTITIDIKDSYTDLLLVGAPMYMGTEKEGKGVYVYAL-NOTRPFYQMSLEPIRQT 547
Qy 542 -----EPPQDARFGFANGALPDNLQDGFADVAGAPLEDHOGALYLYH 586
Db 548 CCSSLDKNSCTKENKNEPCGARTAIKAVKDLNLDGFNDIVIGAPLEDHOGAVYIYH 607
Qy 587 TQSGVRPHPAQRIAAASMPHALSYFGORSVDRLDGLDGLDVAVAGAAQAAILLSSRPV 646
Db 608 SGTIREAYAQRIIPSGDGTILKFPQSGIHGMDLNGDGLTDVTIGLGAALFWARDVA 667
Qy 647 HLTPSLEVTPQALISVVQDRCRRGQAVCLTAALCFQVTSRTPGRWDPHQPMTASLDE 706
Db 668 VVKVTMNFENKVNIOKKCHMEGKETVCINATVCPKSKEDTIYEADLQRYVTLDSL 727
Qy 707 WTAGARAFDGSQRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTVPALDNTTK 766
Db 728 LRQISRSFFSGTQERKQVR--NITVRSECTKHSFYMLDKHDFQDSVRITLDP---NLTD 782
```

Qy	767	P--GPVLNEGSPTSIQKLVPFSKDCGPDNECVTDLVLVQNMIDIRGRKAPFVVVRGRRKV	824
Db	783	PENGPVLDDALPNSVHEHIFPAKDCGNKERCISDOLTINVST---TEKSLLIYKSOHDKF	838
Qy	825	LVSTTLLENKENAYNTSLSIIFSRNLIHLASLTP-QRESPIKVECAAPSAHARLCSVGHPV	863
Db	839	NVSLTVNKGDSAYNTRTVVQHSPLNLIIFSIEBIOKDS-----CESNQNIITCRVGYPF	891
Qy	884	FQTGAKVTFLLEFEFSKSSLLSQVFGKLTASSDSLSRNGTLQBNCTAQTSAIYIOVEPHLLF	943
Db	892	LRAGEVTTFKIIQFNTSHLSENAIHLTSATSDSEEPLESINDNEVNIISIPVKEYEVLQF	951
Qy	944	SSSSTLHRYEVHYPGTLP-----VG-----PGPEFKTTL-----	972
Db	952	YSSASEHHISVAANETIPEFINSTEDIGNEVFYTIKRGHFFMPELQLSISFPNLTA	1011
Qy	973	-----RTNNASCIQNLTEPPG-----PPVHPELOHTNRLNGSNTOCQVVC	1015
Db	1012	GYPVLPIGWSSSDNVNCRPSLEDPPGINSKKMTISKSEVILKRGTIQDCSSTCGVATI	1071
Qy	1016	HLGQLAKG-TEVSVGLRLVHNEFFRPAKFSLTVTSTFELGTGEEGVLQLTASRWSES	1074
Db	1072	TCSILPSDSLQVNVSL--LLWKPTFIARHFSLSNLTLRGELKSENS--LTLSSNRKREL	1128
Qy	1075	LLEVVTQ-RRPILISLWILIGSVGLLALLLVFCLWKLQFFAHKKIPEEEKREE	1128
Db	1129	AIQISKDGLPCRVPLWILLSAFAGLLLLLLIALLWKIGFF---KRPLKKKMEK	1180

RESULT 13

US-09-984-130-103

Sequence 103, Application US/09984130

Publication No. US2003005231A1

GENERAL INFORMATION:

APPLICANT: Ni et al.

TITLE OF INVENTION: 12 Human Secreted Proteins

FILE REFERENCE: PF489P2

CURRENT APPLICATION NUMBER: US/09/984,130

CURRENT FILING DATE: 2001-10-29

PRIOR APPLICATION NUMBER: 60/243,792

PRIOR FILING DATE: 2000-10-30

PRIOR APPLICATION NUMBER: 09/836,353

PRIOR FILING DATE: 2001-04-18

PRIOR APPLICATION NUMBER: 60/198,407

PRIOR FILING DATE: 2000-04-19

PRIOR APPLICATION NUMBER: PCT/US99/25031

PRIOR FILING DATE: 1999-10-27

PRIOR APPLICATION NUMBER: 60/105,971

PRIOR FILING DATE: 1998-10-28

NUMBER OF SEQ ID NOS: 149

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 103

LENGTH: 1151

TYPE: PRT

ORGANISM: Homo sapiens

US-09-984-130-103

	Query Match	31.7%;	Score	1881.5;	DB	10;	Length	1151;	
	Best Local Similarity	35.9%;	Pred. No.	3.2e-157;					
	Matches	426;	Conservative	220;	Mismatches	423;	Indels	119; Gaps	23;
Qy	23	FNLDEHHPLRFLPGPPEAEFGYSVLQHVGCGGRWMLVGNAPWDGSGDRRGDGVYRCVPVGGAH	82						
Dd	1	FNVDVKNKSMTFSGPVDEMDFGYTVOOYEENEGKWLLIGSPLVGPQKNKTGDVYKCPVGRGE	60						
Qy	83	NAPCAKHGLGDYQLGNSHSHPAV-----NMHLGMSILLETGDGGPWACAPLWSRACGSSSVF	137						
Dd	61	SLPCVKLDLP-----VNTSIPNVTVEKENMTGF-STLVTPNGGFLACGPLAYRCGHLY	115						
Qy	138	SSGICARVDASFPQGSLAPTAQRCPTVMOVVIVLDGNSITYPNSEVQTFFLRRLVKGKLF	197						
Dd	116	TGTGCSDPSPFPVVNSGIAP-VQECSTOLDIVI VLDGNSITYPNDSVTAFLNDLLKRMDI	174						

Qy	198	DPQIOVGLVQYGESPVHEWSLGDFTKBEHVBAAKNLSRREGREYKTAQAIWVACTEGF	257
Db	175	GPQOTQVIGVQYGENYTHEFNLNKYSSTEVLEVAAKKIVQRGQRTWALTGTDTARKEAF	234
Qy	258	SQSHGGRPEAARLLVVVTDGESHDGSELPAALKAACEAGRVTRYGIAVLGHYLRQRDPSS	317
Db	235	TEARGARGVKVMVIVTDGESHDNRHLKKVIOCEDENIQRSIALGSSYNRGNLSTEK	294
Qy	318	FLREIRTIASDPPERFFNVVTDAAALTDIVDALGDRIFLGEGSHAENESSFGLMSQIGF	377
Db	295	FVEEIKSIASEPTEKHFNVVSDSLAVTIIVKTEGIRIFALEATADQSAASFEMESQTFG	354
Qy	378	STHRLXGDLFCGMVGYADWCGSVLWLEGGHRLFPFPRMALEDEPPALQNHAAVLYGVS	437
Db	355	SAHYSQDWMLGAVGAYDMNGTVMQKASQIIIPRNTFNVESTKKNPEPLASYLYGTIVNS	414
Qy	438	MLLRGGRRLLPLSGAPRPHRGKVIAFOLKXGDGAVRVAQSLQGEQIGSYFGSELCPDTR	497
Db	415	ATASSGDVLYIAGOPRVNHTQVVIYRM-EDGNIKILQTLSGEQIGSYFGSILLTTDIDK	477
Qy	498	DGTTDVLVVAAPMFLGPONKETGRVYVYLVGQOSLLTLQGTLP-----	541
Db	474	DSNTDILLVAPMYMTEKEEQGVVYAL-NOTREYQMSLEPIKOTCCSSRQHSNCTT	532
Qy	542	--EPQODARFGFAMGALPDLNODGPAADVAGAPLEDGHOGALYLYHGTQSGVRPHPAQR	598
Db	533	ENKNEPCGARFGTAIAVKOLNLDGFNDIVIGAPLEDHGGAVIYHSGKTIKEVYQAR	592
Qy	599	IAAASMPHALSYFCGRSDGRDLDDGDLVDVAVGAQAAIILLSRPVHLTPSLVTPQA	658
Db	593	IPSGDGKTLKFFGQSIHGEMDLNGDGLTDVTIGLGAALFWSRDVAVVKVWNPENK	652
Qy	659	ISVVQDCRRRQGEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLEWTAGARAAFDGS	718
Db	653	VNIQKNCHMEGKETVCINATVCFEVLKSKEDTIYEADIQYRVTLDSLRQISRSFSGT	712
Qy	719	GQRLSPRLRLVGNVTCEOIHPHVLDTSDYLRPVALTVTFALDNTTKP--GPVLNKGSP	776
Db	713	QERKVR--NITVRKSECTKHSFVMLDKHPQDSVRITLDF--NLTDPPENGVLDDSLP	767
Qy	777	TSIQKLVFFSKDCPDNECVTDLVLQYNMDIRSGRKAPFVVVGRRRKVLVSTTLNKRKN	836
Db	768	NSVHEYIPFAKDCGKKEKICISLSLHV---ATTEKDLLIVRSQNDKFNVSILTVKNTKDS	823
Qy	837	AYNTSLIIESRNLHLASLPQRESPIKVECAAPSAHARLCSVGHVPVFTQCAKVTLELFE	896
Db	824	AYNTRTIVHSPNLVFSGI---EAIQKDC--ESNHNITCKVGYPLRGEWMTFKILF	877
Qy	897	EFSCSSLLSOVFGKLTASSDSLRNGTLQENTACTSAYIOVEPHLLPSSSTLHRYEVHP	956
Db	878	QFNTSYLMENVTIYLSATSDSEPPETLSDNVNISIIPVKYEVGLQFYSSASEYHISAA	937
Qy	957	YGTLP-----VG-----PGPEFKTTLRTNNASCIVQNITTEPPGPPVH	993
Db	938	NETVPEVINGSTDIGNEINIFYLIRKSGSPMPELKLIS-----FPNMNTSGYPVLY	990
Qy	994	PEELOHTNRLN-----GSNTQCO-----VVRCHLGOLA	1021
Db	991	PTGLSSSENANCRPHIFEDPFSINGKMTTSTDHLKRGITLDCNTCKFATITCNLTS-S	1049
Qy	1022	KGTEVSVGLLRLVHNEPFRRAKFKSLTWVSFELGTBEGSVLQJTEARWSLESLLVQOT	1081
Db	1050	DISQVNSVL--ILWKPTFIKYSFSSNLITRGLERSENAS-LVLSSSNQKRELAIQISKD	1106
Qy	1082	-RPLISLWILIGSVLGGLLALLVFLWKLGPFAHKKIPBEEKREE	1128
Db	1107	GLPGRVPLWVILLISAFAGLLILLMLLILALWKIGFF--KRPLKKKMEK	1151

RESULT 14
US-09-836-353A-103
; Sequence 103, Application US/09836353A

RESULT 14
US-09-836-353A-103
; Sequence 103, Application US/09836353A

```
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: P489P1
; CURRENT APPLICATION NUMBER: US/09/836.353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-103

Query Match      31.7%; Score 1881.5; DB 10; Length 1151;
Best Local Similarity 35.9%; Pred. No. 3.2e-157;
Matches 426; Conservative 220; Mismatches 423; Indels 119; Gaps 23;

Qy 23 FNLDEHPRLFPFGPPEAEFGYSLQHVGGQRMWLVGAPWMDPGSDRRGDYRCPVGGAH 82
Db 1 FNVVDKNSMTFSGPVEDMFGYTVQYENEBGKWLIGSLVLPVQPKNRTGDYKCPVGRGE 60
Qy 83 NAPCAKGLHDYQLGNSHPAV-----NMHLGMSLLETDGDDGGMACAPLWSRACGSSVF 137
Db 61 SLPCVKLDLP-----VNTSIPNVTVEKENMTFG-STLVTPNPNMGFLACGPLYAYRCGHLHY 115
Qy 138 SSGICARVDASFQPGQSLAPTAQCPTYMDVIVLDGNSNIYPWSEVQTFRLRLVGLFI 197
Db 116 ITGICSDVSPTFQVNSIAP-VQECSTQLDIVLDGNSNIYPWDSVTFAPLNDLKKMDI 174
Qy 198 DPEIQIQLVQYSGSPVHENSIGDPRTEKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGF 257
Db 175 GPKQTQVGIQYQGENVTHEFNLNKYSSTEEVLVAKKIVQRGGQTWTALGTDTRKEAF 234
Qy 258 SQSHGGRPEAARLLVVTDGSHGGEELPAALKACEAGRVTRYIAVLGHYLRQRDPSS 317
Db 235 TEARGARRGVKVMVITDGHSDHNRLLKKVIOQCEDENIQRFSAIILGSYNRCNLSTEK 294
Qy 318 FLREIRTIASDPDRFPFNVTDEAALTDIIVDALGDRIFGLEGSHAENESSFGLMSQIGF 377
Db 295 FVEEIKSIASEPTEKHFNFVSDLEALVTIVKTLGERIFALEATADQSAASFEMMSQTGF 354
Qy 378 STHRLKDGILFGMYGAYDWGSGVLWLGCHRLFPFRMALEDEFPFALQNHAAVLYGYSVSS 437
Db 355 SAHYSQDWMLGAVGADWNGTVVMQKASQIIIPNTTFNVESTKNEPLASYLGYTVNS 414
Qy 438 MLLRGRRLLFUSGAPRFRHRKVIAFQLKXGDGAVRVAAQSLQGEIGSYFSGELCPDTR 497
Db 415 ATASSGDVLYIAGQPRYNHTQVIIYRM-EDGNKILQTLGSEQIGSYFSGILTTDIDK 473
Qy 498 DGTITDVLVAAPMLFGQNKETGRVYVYLVCQSSLTTLQTLQP----- 541
Db 474 DSNTDILLVGAPMYMTEKEEGQKVVYVAL-NQTRFYEQMSLEPIKQTCSSSRQHNSCTT 532
Qy 542 ---BPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDGHQCALYLYHGTSQGVPHPAQR 598
Db 533 ENKNEPCGARGTAIAAVKDLNLGDFNDIVGAPLEDHGHGAVIYHSGKTIKREYAR 592
Qy 599 IAAASMPHALSYFGRSVDRLDLDGDDLDVAVAGQAAILLSRPVHLHLPFSLVTPQA 658
Db 593 IPSGGDGKTLFFQGSIHGMDLNGDGLDVTIIGLGAALFWSRDVAVVAVVWTFNFPNK 652
Qy 659 ISVVORDCRRRGQAVCLTAALCFQVTSRTFGRWDHQFYMEFTASLDSEWTAGARAAPGS 718
Db 653 VNIQKNCHEGKETVCINATVCFEVLKSKEDTIYEADLQYRVTLDSLRLQISRSFFSGT 712

; Publication No. US10741600
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1161
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-600-1161

Query Match      31.6%; Score 1874.5; DB 17; Length 1177;
Best Local Similarity 35.9%; Pred. No. 1.4e-156;
Matches 427; Conservative 219; Mismatches 424; Indels 121; Gaps 24;

Qy 20 CSPENLDEHPRLFPFGPPEAEFGYSLQHVGGQRMWLVGAPWMDPGSDRRGDYRCPVG 79
Db 26 CVSFNVDRKNSMTFSGPVEDMFGYTVQYENEBGKWLIGSLVLPVQPKNRTGDYKCPVG 85
Qy 80 GAHNAPCAKGLHDYQLGNSHPAV-----NMHLGMSLLETDGDDGGMACAPLWSRACGS 134
Db 86 RGSLLPCVKLDLP-----VNTSIPNVTVEKENMTFG-STLVTPNPNMGFLACGPLYAYRCGH 140
Qy 135 SVFSSGGICARVDASFQPGQSLAPTAQCPTYMDVIVLDGNSNIYPWSEVQTFRLRLVGLK 194
Db 141 LHYTTGICSDVSPTFQVNSIAP-VQECSTQLDIVLDGNSNIYPWDSVTFAPLNDLLE 199
Qy 195 LFTDPEIQIQLVQYSGSPVHENSIGDPRTEKEEVVRAAKNLSRRREGRTKTAQAIMVACT 254
Db 200 MDGPKQTQVGIQYQGENVTHEFNLNKYSSTEEVLVAKKIVQRGGQTWTALGIDTARK 259
Qy 255 EGFSSQSHGGRPEAARLLVVVTVDGSHDGBELPAALKACEAGRVTRYIAVLGHYLRQRD 314
```

Db 260 EAFTEARGRGVKKVMVITVDGESHNDHRLKVKIQDCEDENIQRFSAIILGSYNRGNLS 319
Qy 315 PSSFLREIRTASDPDERFFNVNDEAALTDIVDALGDRIFGLEGSHAENESSFGLEMSQ 374
Db 320 TEKFEVEIKSIASEPTEKHPFNVDALAVITVKTLGERIPALEBATAQSAASEMEMSQ 379
Qy 375 IGFSFTHRLKOGILSGMVGAYDMGSGVLWLEGGHRLFPFRMALEDEFFPALQNHAAVILGYS 434
Db 380 TGPSAHYSQDMVMLGAVGAYDMGTVVMQKASQIIIPRNTTFNVESTKKNPLASYLYGT 439
Qy 435 VSSMLLRGGRLLFSLGAPFRFRHRGKVIAPQLKDKGAVRVAQSLQCEQIGSYFGSELCPLD 494
Db 440 VNSATASGDVLYIAGQPRYNHTQVLIYRM-EDGNIKILQTLGEGIGSYFGSILITTD 498
Qy 495 TDRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOSSLTLQGTLP----- 541
Db 499 IDKDSNTDILLVGA PMVGMTEKEGKVYVAL-NQTRFEYQMSLEPIKQTCSSRQHNS 557
Qy 542 -----EPPQDARFGFANGALPDINODGFADVAVGAPLEDHOGALYLYHGTQSGVRPH 595
Db 558 CTTENKNEPCGAREGTAIAVKDNLNODGFNDIVIGAPLEDHGGAVIYHSGGKTIKEY 617
Qy 596 AQRIAAASMPHALSYFGRSVDRGLDLDGDDLDVAVGAQGAAILSSRPVHLTPSLEVT 655
Db 618 AQRIPSGCGDKTLAFEGSGIHGMDLNGDGLTDVTIGGLGAALFWSRDVAVVKVWNFE 677
Qy 656 POAISVQVORDCRRGQEAVALCFOVTSRTFGRMDHOFYMRFTASLDEWTAGARAAF 715
Db 678 PNKVINIKKNCHMEGKETVCINATVCFDVKLKSKEDTIYEADLQVRVTLDSLRQISRSFF 737
Qy 716 DGSQRLSPRLRLSVGNVTCEQHLHFVLDTSYLRPVVALTVFALDNTTKP--GPVLNE 773
Db 738 SGTQERKQVR--NITVRKSECTKHSFYML--HDFQDSVRITLDF---NLDPENGPVLDD 790
Qy 774 GSPTSIOKLVFPCDGPDCVTDVLQVNMIDIRGSRKAPFVVRGGRKVLVSTLTENR 833
Db 791 SLPNSVHEYIIFAKDCGKKEKISDLSLV---ATTEKDLIIVRSQNDKFNVSILTAKNT 846
Qy 834 KENAYNTSLSIIFSRNLHSLTPQRESPIKVECAAPSAAHARLCSVGHVPVQTGAKVTL 893
Db 847 KDSAYNTRTI VHYSPNLVFSGI---EAIQKDCS--ESNHNITCKVGYFFLRGRGEMVTFK 900
Qy 894 LEPEFSCSLLSQVFGKLTASSDSLRNGTLQENTAQTSAYIOYEPHLLFSSESTLHRYE 953
Db 901 ILFOFNTSYLMENVTIYLSATSDSEPPETLSDNVVNISIPVKYEVGLQFYSSASEYHIS 960
Qy 954 VHPYGTLP-----VG-----PGPEFKTLRTNNASCIQNLTEPPGP 990
Db 961 IAAETVPEVINSTEDIGNEINI FYLIRKSGSFPMPELKLIS-----FPNMTSNGYP 1013
Qy 991 PVHPEELQHTNRLN-----GSNTQCO-----VVRCHLG 1018
Db 1014 VLYPTGLSSSENANCRPHI PEDPPSINSNGKMTTSTDHLKRGTLDCNTCKFATITCNLT 1073
Qy 1019 QLAGKTEVSGVGLRLVHNHFFRRAKFKSLTVVSTFELCTEGSVLQLTEASRWSESLELV 1078
Db 1074 S-SDISQVNVSL--ILWKPTFKSYFSSNLNLTIRGELSENAS-LVLSSNQKRELAIQI 1129
Qy 1079 VQT-RPILISLWILIGSVLGLLALLALVFCMLKLGPFHKKIPEEKREE 1128
Db 1130 SKDGLPGRVPLWILLAFAGLLMLLILALWKIGFF---KRPLKKONEK 1177

Search completed: April 6, 2005, 12:53:09
Job time : 148.389 secs

This Page Blank (uspto)

403	LGYSTELALWKGVQNLVL-GAPRYQHTKAVITQVSRQW--RKXAEVGTGTQIGSYFGAS	455
490	LCPLDTRDGTDTDLVLAAPMFLGPQNKETGRVYVYL--GOOSLLTLTGTLQPPPPQD-	546
460	LCSVDVDSGSDTLILIGAPHYY--BQTRGGQVSVCPPLPRGQVQWQCDVLRGEGHPW	517
547	ARFGFAMGALPDNLQDGFADVAVGAPLEDHOGALYLYHG-TQSQVVRPHAPQRIAAAMP	605
518	GRFGAALTVLGDVNEKLDIVATGAFGEQENRGAVALPHGASGESGTPSHSORIASQLS	577
606	HALSIFGRSVDRGLDLDGDDLVNVAAGAAIILSSRPVILHVTLSLETPVQPAISVQRD	665
578	PRLOYFGQALSGGGDLTQDGLMDLAVGARGQVLLLSLPLVKGVAMRSPVEVAKAVR	637
666	C-----RRRGOEAVCLTAALCFQVTSRTPGRWDHQFMYRMTASIDETAGARAAFDGS	718
638	CWEEKPSALEAGDATVCLTIQ-----KSSLDQLGDIOSSVRFDLALDPCRLTSRAIFNET	692
719	GORLSPPRLRLSVGNVTCQLHPIHVL-DTSYLRPVVALVTTFALDNTTKPG-----PVUN	772
693	KNPTLRRKRTGLG-ICETLKLPLDPCVEDVVSPIILHNLFSRLVREPISQNLRPVLA	751
773	EGSPTSIOKLVPESKDCGPDNECVTLVLQVMDIRGSRKAPVVRGGRKVLVSTLLEN	832
752	VGSQDLFTASLPPEKNCQDGLCEGD--LGVTLSPFGLQD--LTVGSLELNLVIVTVMN	806
833	RKENAYNTSLIIFSRNLHLASLTPQRESF---IKVEC-AAPSAHARL---CSVGHPV	883
807	AGEDSVGTVVSLYYPAGLSHRRVSGAKQPHOSALARLACETVPTDEGLRSSRCVNHPI	866
884	FOTGAKVTFLLPEFSCSLLSQVFGKLTASSDLSLERNGTLOQNTAQTSAIYQYEPHLLF	943
867	FHEGNGTFTVTPDVSVKATLGDRM-LMRASASSENKASSKATFQLELPVKYAVYTM	925
944	S--SESTL-----HRYEVHPGTLPVGPGEPEKFTLRLNNASCIVONIT	985
926	SROEESTKYNFATSDKKEAHRHYRVNNLSQORDLAITSINFVFWVLLNGVA-VMDVVM	984
986	EPPG-----PPVHPPELQHTNR---LNGSNTQCOVVRCHLGQAKGTEVSVGLLR	1033
985	EAPSQSLPCVSEKRPQHSDFLTQISRSPLMDCSTADCLQFCDVPSVQELDFTLKG	1043
1033	LVNEFFRRAKFKSLTVSTFELGTGEEGVSQVLOLTASRWSESLLE-VVOTRPIILSLMIL	1099
1045	NLSFGWVRETLOQKVLWSVAEITFDTSVYSQLPQGEAFMRAQMEWLEEDSVYNAIPII	1103
1092	IGSVGLGILLALLVCLMKLGF-AHKKIPEEKREE	1128
1105	MGSSVALLLALITATLYKLGFFRKHYLEMDKPED	1142

RESULT 13

US-08-482-293A-2

Sequence 2, Application US/08482293A

Patent No. 5831029

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sear Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

Db	STREET: 233 South Wacker Drive, 6300 Sear Tower	
Qy	CITY: Chicago	
Db	STATE: Illinois	
Qy	COUNTRY: United States	
Db	ZIP: 60606-6402	
Qy	COMPUTER READABLE FORM:	
Db	MEDIUM TYPE: Floppy disk	
Qy	COMPUTER: IBM PC compatible	
Db	OPERATING SYSTEM: PC-DOS/MS-DOS	
Qy	SOFTWARE: PatentIn Release #1.0, Version #1.25	
Db	CURRENT APPLICATION DATA:	
Qy	APPLICATION NUMBER: US/08/605,672	
Db	FILING DATE: US/08/605,672	
Qy	CLASSIFICATION: 530	
Db	PRIOR APPLICATION DATA:	
Qy	APPLICATION NUMBER: US 08/173,497	
Db	FILING DATE: 23-DEC-1993	
Qy	PRIOR APPLICATION DATA:	
Db	APPLICATION NUMBER: US 08/286,889	
Qy	FILING DATE: 5-AUG-1994	
Db	PRIOR APPLICATION DATA:	
Qy	APPLICATION NUMBER: US 08/362,652	
Db	FILING DATE: 21-DEC-1994	
Qy	ATTORNEY/AGENT INFORMATION:	
Db	NAME: Williams Jr., Joseph A.	
Qy	REGISTRATION NUMBER: 38,659	
Db	REFERENCE/DOCKET NUMBER: 27866/32684	
Qy	TELECOMMUNICATION INFORMATION:	
Db	TELEPHONE: 312-474-6300	
Qy	TELEFAX: 312-474-0448	
Db	TELEX: 25-3856	
Qy	INFORMATION FOR SEQ ID NO: 2:	
Db	SEQUENCE CHARACTERISTICS:	
Qy	LENGTH: 1161 amino acids	
Db	TYPE: amino acid	
Qy	TOPOLOGY: linear	
Db	MOLECULE TYPE: protein	
Qy	US-08-605-672-2	

Query Match 19.0%; Score 1126; DB 2; Length 1161;

Best Local Similarity 29.5%; Pred. No. 7.5e-99;

Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;

Qy	23 FNLDHHPRLFPQGPBAEFGYSVLQHVGGQRMVLVGAIPWDGPGSGDRGDVYRCVPVGAH	82
Db	17 FNLDVEEPTIFQ-EDAGGFGQSVQF--GGSRLVVGAPLEVVAAQTGLRYDC---AAA	69
Qy	83 NARCAGHLGDYOLGNSHIPAVNMHLGMSLLETDDGGGFMACAPLWSRACGSSVFSSGIC	142
Db	70 TGMCOPIPL-----HIRPEAVNMSLGLTAASTNGSRLLAGCPTLHRVCGENSYSKGC	123
Qy	143 ARVDASGFQPGSLAPTAQCP--TYMDVWTVLDGNSI--YPMSEVQTFRLRLVGLKFLDTP	199
Db	124 LLGSRWEIIQTVPDTPPCPHQEMDIVLIDGSGIDQNDQNMKGFFVQAVMGQ--FEG	181
Qy	200 EQIQVLQVQGESPVHWSLGFRTKEEVVRAAKNLSRREGRETAKAIWVACTGEGFSQ	259
Db	182 TDTLFAIMQVSNLTKHTFTQRTSPSQSLVDPIVQLKGL-TFTATGILTVVTQLFHH	240
Qy	260 SHGGRPEARLLVVVTDGESH-DGELPALKACAGRVTRVGIAGVHYLRQRDPDSF	318
Db	241 KNGARKSAKKILIVTDGQKDPLEYSVPIQAEKAGIIRYAGV-GHAF---QGPTA-	295
Qy	319 LREIRTIASDPDERFFNTVDEAALTDIVDALGDRIFGLGEGSHAENESSFGLMSOIGFS	378
Db	296 RQELNTISSAPQDHVKVDNFALAGSIQKQLEKIYAVEGTSRASSSFQHEMSQEGFS	355
Qy	379 THRLKDGILFGMVGAYDWGCVSLWLEGGHRLPPRMALEDEFPALQNH-----AY	430
Db	356 TALTMGDLFLGAVGSFSW-----SGGAPLYPPNMS-----PTFINNSQENVMDRDSY	402
Qy	431 LGYSVSMMLRGRRRLFLSGAPRFRHGKVIAP-QLKDGAVRVQAQSLQGEIGSYFSGE	489

```
; APPLICATION NUMBER: US/08/482,293A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-482-293A-2

Query Match 19.08; Score 1126; DB 2; Length 1161;
Best Local Similarity 29.54; Pred. No. 7.5e-99;
Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;

QY 23 FNLDEHHPLRPPGPEAEFGVSLQHVGGORWMLVGPAPWDGPGDRRGDVYRCPVGGAH 82
DB 17 FNLDBVEPTIFQ-EDAGGFQGVVQF--GSGR-LVVGAPLEVVAANTGLDYC---AAA 69
QY 83 NAPCAKGLGDYQLGNSHSPAVNMHQLMSLLETDGCGFMACAPLWBRACSSVFFSGIC 142
DB 70 TGMCPQIPPL-----HIREAVNMSLGLTLAASTNGSRLLACGPTLHRVCGENSYSKGC 123
QY 143 ARVDASPOGSLAPTAQRCP-TYMDVVIVLDGNSI--YPMSEVQTFRLRLVKLFIDP 199
DB 124 LLLGSRWEIQTVPDAPTECPHQEMDIVFLIDGSGSIDQNDFMQKGFVQAVMQ--FEG 181
QY 200 EQIQVGLVQYGESVPVHEWSLGDFTKEEVVRAAKNLRRREGRETAKTAQAIMVACTEGFSQ 259
DB 182 TDTLPALMQVSNLLKHIFTTQFTSPSQSLVDPIVQLKGL-TFTATGILTVVTQLFHH 240
QY 260 SHGRPRPAAALLVVVTDGESH-DGEELPAALKACEAGRVTRYGIYAVLGHYLRQRDPSSF 318
DB 241 KNGARKSAKKILIVITDGQYKDPLEYSDVIPQAEKAGIIRYAGV-GHAF---QGFTA- 295
QY 319 LREIRTTASDPDERFFNVYDDEALTDIVDALGDRIFGLEGSHAENESSFGLEMSQIGFS 378
DB 296 RQELNTISSAPPQHVDFKVDNFALGSIQKQLQEKIYAVEGTOSRASSSFQHEMSQEGFS 355
QY 379 THRLKQIGLFGWGVGAYDWGSLVLEGGHRLFPPRMALEDEFPFALONHA-----AY 430
DB 356 TALTMQGLFLGAVGSFVW-----SGAFLYPPNMS-----PTFLNMQENVDMRDSY 402
QY 431 LGYSVSMLLRGRRLLFSLGAPFRHRKGVIAF-QLKKGAVRVAQSLQGEQIGSYFGSE 489
DB 403 LGYSTELALWKGVQNLVL-GAPRYQHTGKAVIFTQVSRQW--RKAETVGTQIGSYFGAS 459
QY 490 LCPDLTDQDGTDDVLLVAAPMFLGPQNKETGRVVYLV--GQQSLLLTQGLTQPEPPQD- 546
DB 460 LCSVDVDSGDSGTDLLIGAPHYY--EQTGCGQSVCPCLPRGQRVQWQCDVLRGEQHPW 517
QY 547 ARFGFAMGALPDNLQDGFADVAVCAPLEDHGQALYLVHG-TQSGVPHPAQRTAAASMP 605
DB 518 GRFGAALTIVGDVNNEDKLDVAIGAPGEQENRGAVYLFHGAESGIFSFSHQRIASSQLS 577
QY 606 HALSYFORSVDGRDLDDGDDLDLDVAVGAQGAAILLSRRPIVHLTPPSLEVTPQAISVVQRD 665
```

```
DB 578 PRLQYFQALSGGQDLTQDGLMDLAVGARGQVLLRLSPVLKVGVMRSPFVAVKAVYR 637
QY 666 C-----RRRQGEAVCLTAALCFQVTSRTPTGWDHQFYMRFTASLDDEWTAGARAPFDGS 718
DB 638 CWEEKPSALEAGDATVCLTIQ-----KSSLDQLGDIQSSVRFDLALDPGLRTSALPNET 692
QY 719 GQRLSPRRRLSLVGNVTCBQLHFHVLDTSDYLRPVALTVTFALDNTTKPG-----PVLN 772
DB 693 KNPTLTKRKTGLG-ICETLKLPLDCVEDVVSPILHLNPSLVREPISPQNLRPVLA 751
QY 773 EGSFTSIQKLVPFSKCGPDNECVTLVLQVNMDIRSKAPFVVRGGRKVLVSTTLN 832
DB 752 VGSQDLFTASLPPFKNCGQDGLCEGD--LGVTLSFSGLOQT--LTVGSSLELNVIVTVN 806
QY 833 RKENAYNTSLIIFSRNLHLASLTPORESP-----IKVEC-AAPSAHARL-----CSVGHPV 883
DB 807 AGEDSYGTVVSLYYPAGLSHRRVGAQKQHQHQAALRACVTVEDEGLSSRCNVNHP 866
QY 884 FQTGAKVTFLLFEFSCSSLLSQVFGKLTASSDSLERNGLTQENTTAQTSAYIQYEPHLF 943
DB 867 FHGSGNGTFTVFDVSKATLGDRM-LMRASASSENKASSKATQLELPVKYAVVTMI 925
QY 944 S--SESTL-----HRYEVHPYGTLPVGPQPEPKTTLRTNASCIVONLT 985
DB 926 SROESTKYFNFATSDKKMKAEHRYRVNMLSQRDLAISINFWPVLLNGVA-VMDVVM 984
QY 986 EPPG-----PPVHPEELOHTNR---LNGSNTQCQVVRCHLGLQAKGTESVGLLR 1032
DB 985 EAPSQSLPCVSEKRPQHSDFLTQISRPMLDCSIADCLQRCDFPSFVQEEUDFTLKG 1044
QY 1033 LVNEFFRAKPKSLTVSTFELGTBEGSVLQLTEASRWSESLLE-VVQTRPILISLWIL 1091
DB 1045 NLSFGWVRETLQKKLVVSAEITFTDSVYSQLPQGEAFMRAQMWLEDEVVNAIPIL 1104
QY 1092 IGSVGLGILLALLVCLMKLPGFF-AHKKIPEREKREE 1128
DB 1105 MGSVGALLLILATITLVKLGFFKRYHEMLDKPED 1142.

RESULT 14
US-08-943-363-2
; Sequence 2, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
```

Qy	773	EGSPTSIOKLVPSCKDPPNECVTLDLVQVNDMDIGSRKAPFVWVGRRKVLVSTTLEN	832
Db	752	VGSDQLFTASLPPEKNCQGDLCEGD--LGVTLFSGLQT--LTVGSSLELNVIVTVMN	806
Qy	833	RKNAYNTSLSIIFSRNLHLASLTPORESP----IKVEC-AAPSAUARL----CSVGHPV	883
Db	807	AGEDSTGVTVSLYYPAGLSHRRVSGAQKQPHQSALHLACETVTEDEGLSSRCNVNHP	866
Qy	884	FOTGAKVTFLLPEFSCSSLLSQVFKLTASSDLSLERNGTLOBNTAQTSAIYIYEPHLLF	943
Db	867	FHEGSGNGTFVTFDVSFKATLGDRM-LMRASSASSENKASSKATQLELPVKVAVVTMI	925
Qy	944	S--SESTL-----HRYEVHPGYCTLVPGPGPEFKTILRTNNASCIQVNL	985
Db	926	SROEESTKYFNFASTDSEKMKKEAHRVYRNNLSQORDLAISINFWPVLNGLVA-VMDVVM	984
Qy	986	EPG-----PPVHPEELQHTNR---LNGSNTOCQVVRCHLGLQAKGTVEVSGLLR	1032
Db	985	EAPSQSLPCVSEKRPQHSDFLTQISRPMLDCSIADCLQFRCDVPSFSVQEBLDTLKG	1044
Qy	1033	LVHNEFFRAKFKSLTVVSTFELGTBEGSVLQLTEARMSSESILLE-VVQTRPILISLWIL	1091
Db	1045	NLSFGWVRETQKKVLVSVSAEITFDTSVYSQLPQGEAFMRAQWEMVLEDEVVNAIP	1104
Qy	1092	IGSVLGGLLLLALLVCLWKLGFF-AHKKIPPEEKREE	1128
Db	1105	MGSSVGALLLALITATLYKLGFFKRYKEMLEDKPED	1142
RESULT 15			
US-09-193-043-2			
; Sequence 2, Application US/09193043			
; Patent No. 6251395			
; GENERAL INFORMATION:			
; APPLICANT: Gallatin, Michael W			
; APPLICANT: Van der Vieren, Monica			
; TITLE OF INVENTION: No. 6251395el Human 2			
; FILE REFERENCE: 27866/35004			
; CURRENT APPLICATION NUMBER: US/09/193,043			
; CURRENT FILING DATE: 1998-11-16			
; EARLIER FILING DATE: 1993-12-23			
; EARLIER FILING DATE: 1993-12-23			
; EARLIER FILING DATE: 1994-08-05			
; EARLIER FILING DATE: 1994-08-05			
; EARLIER FILING DATE: 1994-12-21			
; EARLIER FILING DATE: 1994-12-21			
; EARLIER FILING DATE: 1997-10-03			
; NUMBER OF SEQ ID NOS: 114			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 2			
; LENGTH: 1161			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-193-043-2			
Query Match 19.0%; Score 1126; DB 3; Length 1161;			
Best Local Similarity 29.5%; Pred. No. 7 5e-99;			
Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;			
Qy	23	FNLDHHPRLPFGPPEAFBFCYSVLQHVGGQRMVLVGAPWDGPGSDRRGDVYRCPVGGAH	82
Db	17	FNLDVEPTIFQ-EDAGGFGQSVQF--GGSR-LVVGAPLEVVAANQTRLGYDC--AAA	69
Qy	83	NAPCAKHLGDYQLGNSHSPAVNMHLGMSLLETDGGGFACAPLWSRACGSSVFSGIC	142
Db	70	TGMCCQPIPL-----HTRPEAVNMSLGLTLAASTNGSRLLACGPTLHRVCGENSYSKGC	123
Qy	143	ARVDASPOGSLAPTAQRCP-TYMDVVIVLDCSNSI--YPWSEVQVFLRLVGLKLPIDP	199
Db	124	LLLGSRWEIITQVDPDATPECPHOEMDIVFLIDGSGSIDQDNDFNMKGFGVAQVMQ--FEG	181
Qy	200	EOIQOVLVOYGESPVHSEWSLGDFTKEEVVRAAKNLSRRREGRETKTAAQIMVACTEGFSQ	259

```
Db      182  TDTLFAIMQSNLKHIFTFTQRTSPSQSLVDPIVOLKGL-TFTATGILTVVTQLFHH 240
Qy      260  SHGGRPEAARLLVVVTGESH-DGEELPAALKACEAGRVTRYGIATVIGHYLRQRORDPSSP 318
Db      241  KNGARKSAKKILVITDQKIDPLEYSDVIPQAEKAGIIRIYAGV-GHAF---QGPTA- 295
Qy      319  LREIRTIASPPDERFFNFVDEAALTDIVDALGDRIFGLEGSHAENESSFQLEMSQIGFS 378
Db      296  RQELNTISSAPPQDHVPKVDNFAALGSIQKOLQEKIYAVEGTQSRASSSQHEMSQEGFS 355
Qy      379  THRLKOGILFGMVGAYDWGGSVLMLEGGHRLFPPRMALEDEFFPALQNH-----AY 430
Db      356  TALTMDGLFLGAVGSFSW-----SGGAFLYPPNMS-----PTFINMQENVDMRDSY 402
Qy      431  LGYSVSSMLRGRRRLFLSGAPRRHREKVIAP-OLKKDGAVRVAQSLQGEQIGSYEGSE 489
Db      403  LGYSTELALMGVQNLVL-GAPRYQHTGKAVIFTQVSRQW--RKKAETGTQIGSYFGAS 459
Qy      490  LCPLDTRDGTDTVLLVAAMPFLGPQNKETGRVYVYL--GQSLLLTLQGTLOPEPPQD- 546
Db      460  LCSVDVDSGSTDLILGAPHY--EOTRGGQSVCPPLRGQVQWQCDVLRGEQGHWP 517
Qy      547  ARFGFANGALPDLNODGFADVAGAPLEDGHQGALYLYHG-TQSGVRPHPAQRIAAASMP 605
Db      518  GRFGAALTVLGDVNEDKLIDVAIGAPGEQENRGAVYLFHGASESGISPSHSQRIASSQLS 577
Qy      606  HALSYFGRSVDRLDLDGDDLDVDVAGNAGNAIILSSRPVHLTPSLEVTPOAISVVQRD 665
Db      578  PRLQYFOALSGGGQDLTQDGLMDLAVGARGQVLLRLSLPVLKGVAMRFSPEVAKAVYR 637
Qy      666  C-----RRRQBAVCLTAALCFQVTSRTPGRWDHQPYMRFTASLDEWTAGARAAFDGS 718
Db      638  CWEKPSALEAGDATVCLTIQ-----KSLDLQLDGIQSSVRFDLALDPGLTSAIFNET 692
Qy      719  GQRLSPRLRLSVGNVTCQLHFHVL-D-TSYLRPVALTVTFALDNTTKPG-----PVLN 772
Db      693  KNPTLTRRKTGLG- IHCETLKLPLDCVEDVVSPIILHLNFSLVREPIPSQNLRPVLA 751
Qy      773  EGSPTSQKLVPPFSKDCPDNECVTDVLQYNMDIRGSRKAPFVVRGRRKVLVSTTLEN 832
Db      752  VGSQDLFTASLPFFKNCQDGLCEGD--LGVTLFSGLT---LTVGSSLELNVIVTWN 806
Qy      833  RKENAYNTSLSIIFSRNLHLASLTPQRESP----IKVEC-AAPSAHARL-----CSVGHPV 883
Db      807  AGEDSYGTVVSLYYPAGLSHRVSGAQKQPHQSALRLACEIVPTDEGLRSSRCSVNHPI 866
Qy      884  PQTGAKVTFLLBFFSCSSLLSQVFGKLTASSDSLERNGTLQENTAOQTSAYIQYEPHLLF 943
Db      867  PHEGSGTFTVTFDVSYSKATLGDRM-LMRASASSENNKASSKATFQLELPVKYAVVTMI 925
Qy      944  S--SESL-----HRYEVHPYGTLPVGPPEPKTTLRTNNASCIVQNL 985
Db      926  SRQESTKYFNATSDKKMKAEHRYRVNNSQLDLAISINFVWPVLLNGVA-VMDVVM 984
Qy      986  BPPG-----PPVHPELQTNR---LNGNTOCQVVRCHLQOLAKGTEVSVGLLR 1032
Db      985  EAPQSILPCVSRERPPQHSDFLTQISRSPMLDCSIADCLQRCDDVPSPVQEEIDFTLKG 1044
Qy      1033  LVHNEFFRRAKFKSLTVVSTPELTGTEGSLVQLTEAGRWSSLEL-VVQTRPILISLWIL 1091
Db      1045  NLSFGWVRRETLQKKVLVVSVAEITFTDSVYSQLPQEAQFMRQAQMEMVLEEDVYNAIPII 1104
Qy      1092  IGSVLGGLLLALLVFLWKLGF- AHKKIPEEKREE 1128
Db      1105  MGSSVGALLLALITATLYKLIGFKRHYKEMLEKPED 1142
```

Search completed: April 6, 2005, 12:24:46
Job time : 72.3117 secs

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6189
; LENGTH: 1181
; TYPE: PR1
; ORGANISM: Human
US-09-949-016-6189

Query Match          29.2%; Score 1733.5; DB 4; Length 1181;
Best Local Similarity 33.7%; Pred. No. 1.6e-157;
Matches 410; Conservative 217; Mismatches 431; Indels 159; Gaps 32;

Qy 11 LPLVFLTGL-----CSPFNLDHHPRLFPGPPEAEFGYSLVQHVGQGRWMLVGAPWD 63
Db 11 LPLLLVLAISQILNCLAYNVGLPEAKIFSGPSEQFGYAVQVQFIPKGNWLLVGSFWS 70

Qy 64 GPSGDRGDVYRCVPVGAHNAKAGHL-GDYQLGNSHPAVNMHLGMSLLETDCGGFM 122
Db 71 GFPENRMGVYKCPV-DLSTATCEKLNQTSIPNVTMKTNNLSGLILTRNMTGGFL 129

Qy 123 ACAPLWSRACSSVFSGGICARVDASPOQSLAPTAQRCPTMYDMVIVLDGNSIYPWS 182
Db 130 TCGFLMAQQCGNQYITTTGVCSDISPDFQLSASFSPATQPCPSLDIVVWVCDENSIYPWD 189

Qy 183 EVOTFLRLVGLKLFIDPEIQVGLVQVGPVHWSLGDPRTKBEVVRAAKNLSRRGRE 242
Db 190 AVKNFLEKFOGLDIGNTKVQGLIQVANNPRVFNLTNTYKTEEMIVATSQTSQYGGDL 249

Qy 243 TKTAQAIMVACTEGFSQSHGGRPEARLLVVVTDGESHGDBELPAALKACBAGRVTYGI 302
Db 250 TINTFGAIQYARKYAYSAAAGRRSATKVMVVVTDGESHGSMKAKVIDQCNDHNLRFGI 309

Qy 303 AVLGHYLRORDPSPFLREIRTIASDPDERFFNVNVTDEAALTDIVDALGDRIFGLESHA 362
Db 310 AVLGYLRNALDTKNLKEIKAIASIPTEYFFNVNVDAALEKAGTLEQIFSIETG-V 368

Qy 363 ENESFGLMSQIGFST--HRLKDGILFGMGVADWGSVLW-LEGGHRLPPPRMALEDE 419
Db 369 QGGNFQWMSQVGSADYSSQNDILMLGAVGAFGWSGTIVQKTSHGHLIFP-----KQA 423

Qy 420 FPPALQ--NHAAYLGYSVSSMLRGGRLFLSGAPRFRHRGKVIAFOLKDGAVRVQSL 477
Db 424 FDQILQDRNHSSYLGSVAA-ISTGESHFVAGAPRANYTQGIVLVSVNENGNTIVQAH 482

Qy 478 QGEQIGSYFGSELCPDTRDGTGTVLLVAAAPMLGPONKETGRVYV-----LVGQSL 532
Db 483 RGDQIGSYFGSVLCSDVDKDTITDVLVAGAPMTMSDLKKEGRVYLFITKKGILGQHOF 542

Qy 533 LTLOGTLQPEPPQDARFGFAGALPDNLQDGFADVAVGAPLEDHGOALYLYHGTOSGRV 592
Db 543 --LEG---PEGIENTRFGSAALASDINMGDFNDVIVGSPLENQNSGAVIYNGHQGTIR 597

Qy 593 PHPAQRIAAA--SMPHALSYFGSRVDGRDLDDGDLVDVAVGAQCAAILSSRRPIVHLTP 650
Db 598 TKYSQKILGSDGARFSLHQLYFGRLDGYDGLNGSDITDVSIGAGQVQVQLWSQSIADVAI 657

Qy 651 SLEVTPOAISVVQRDCRRRQGEAVCLTAALCFQVTSRTPGGRDHFQYNNRFTASLDWTAG 710
Db 658 EASFTPEKITLVNNAQ-----IILKLCFSAKFR-PTKQNNQVAIVNITLDA----- 704

Qy 711 ARAAFDGGQRLSPRL-----RLSGVNV-----TCEQLHFHVLDTSDYLRFPVALTVT 758
Db 705 -----DGFSSRVTSRGLPKENNERCLQKNMVVNAQSCPEHIIYIQEPSDVVNSLDLRVD 759

Qy 759 FALDNTTKPG--PVLNIGSPSTSIQKLVFPFSKDCPDNECVTDLVLQVNMIDIRSGKAPFV 816
Db 760 ISLEN--PGTSPALAEVSETAKVFSIPFKDCGEDGLCISDLVDVR-QIPAAQEQPFI 815

Qy 817 VRGGRKVLVSTTLNKRKENAYNTSLSTIFSRNLHLASLTPORESPI---KVEC-AAPSA 872
Db 816 VSNQNKRLTFVTLKNRKESAYNTGI VVDFSENLFASFS---LPVDGTETVTCQVAASQ 871

Qy 873 HARLCSGHVPFQGAUKATFLEPEFSFCSLLSQVFGKLTASSDSLEBNGTLQENTAGTS 932
Db 873 HARLGHYLRORDPSPFLREIRTIASDPDERFFNVNVTDEAALTDIVDALGDRIFGLESHA 362
```


532	Qy	LLTLQGTLPBPQDARFGFAMGALPDLNODGFADVAVGAPLEBHGOGALYLHGTQSGV	591
542	Db	F--LEG---PEGIENTFGSAIALSINDMDFNDVIVGSPLEQNQSGAVIYNGHGTI	596
592	Qy	RHPAQRAAAA--SMPHALSVFGRSDVGRLLDLOGDDLVDVAVGAQGAAILLSRPIVHLT	649
597	Db	RTKYSQKILGSDGAFRSHLQYFGRSLDYGDLNGDSITDVSGIAGFQVVQLWQSGIADVA	656
650	Qy	PSLEVTPOALISVVQDRCRRRQGEAVCLUTAAALCFQVTSRTPGRWMDHQPYMRFTASLBSWTA	709
657	Db	IEASFTEPKITLVNKNQAQ-----IILKLCFSAKFR-PTKQNNQVAIVYNTLDA---	704
710	Qy	GARAAAPGSGQRSLSPRRL-----RLSVGNV-----TCBQLBPHVLDTSDYLRPV--ALT	756
705	Db	-----DGFSSRVTSRGLPKENNERCLQKMWVYNAQSCPEHIYIQEPDSVNVSLDDLRL	758
757	Qy	VTFALDNTWTKPG--PVLNEGSPTSIQKLVPFSKDCGPDNBEVCVTDLVILQVNMDIRGSRKAP	814
759	Db	VDISLEN-----PGTSPALAEAYSETAKVPSIPFHKDCGDKGICISDLVLQ--DVRIPAAQEQP	814
815	Qy	FVVRGGRKVLVSTLTLENRKENAVNTSLSIIFSLENLHASITPQRESPI---KVEC--AAP	870
815	Db	FIYVSNQMKRITFVTLKNKRRESAYNTGIVDFSENLFASFS-----LPVQGTVEVTCVAA	870
871	Qy	SAHARLCSVGHVPFQTKAKVFTLLEFPFSCSSLLSQVFGKLTASSDSLERNGTLQENTAQ	930
871	Db	SQKSVACDVGVYPALKRSEQQVTFITNFDNLQNLQNAQSLSQALSESQEEENKA--DNLVN	928
931	Qy	TSAYIQVPEHLLPSSSSTLHRYEVHPYGTLP-----VGPGEPEKTTLRTNNASCIVQ--	982
929	Db	LKIFLLYDABAIHLTRSTNINFIYEISSDGNVPSIVHSPFEDVGPKFIFSLKVTTVGSPVMSA	988
983	Qy	-----NLTEPPGP-----PVHP-----BELOHTNR	1002
989	Db	TVIIHIPOYTKERNPLMYLTGVQTDKAGDICCNADINPLKIGTSSSVSPKSENFRTKE	1048
1003	Qy	LNGSNTOCQVVRCHLGOLAKGTEVSVGLLRLVHNHFFRRRAFKSLTVVSTFELGTEEGSV	1062
1049	Db	LNCRTASCNNVTCWLKDVHMKGEYFVNVTTRIMNGTFPASSTFQTVQLTAAAEINTYNPEI	1108
1063	Qy	LQLTEASRWSSELLEVQTRPILLSLMI-----LIGSVGLGILLILALVFL	1109
1109	Db	Y-----VIEDNTVTIPLMIMKPDKEKAEVPTGVIIGSITAGILLALLVAIL	1154
1110	Qy	WKLGFFAHKXIPBEEKREKLEQ	1132
1155	Db	WKLGFYKRYEKQWTKKNPDEIDE	1177

RESULT 8

```

US-08-173-497-2
; Sequence 2, Application US/08173497
; Patent No. 5437958
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van Der Vieren, Monica
; TITLE OF INVENTION: No. 5437958e1 Human 2 Integrin Alpha
; TITLE OF INVENTION: Subunit
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 S. Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

Db 693 KNPTLTRKTLGLG-IHCETLKLLPDCVEDVSPPIILHLNFSLVREPIPSQNLRPVLA 751
Qy 773 EGSPTSQKLVPSKDCGPONECVTDLVQVNMMDIRSKAPFVVRGGRKVLVSTLEN 832
Db 752 VGSQDLFTASLPKCNQDGLCEGD--LGVTLSFSGLOT---LTVGSSLELNVITVMN 806
Qy 833 RKNAYNTSLSIIFSRNLHLASLTPORESP-----IKVEC-AAPSAHARL-----CSVGHPV 883
Db 807 AGDSYGVTVSVLYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLASSRCSVNHPI 866
Qy 884 FOTGAKVTFLLEPEFSCSSLLSQVFGKLTASSDSLENGTLQENTATSAIYQEPHLLF 943
Db 867 FHESNGTFTVTFDVSYSKATLGDRM-LMRASASSENKASSKATFQLELPVKYAVVTMI 925
Qy 944 S--SESTL-----HRYEVHPYGTLPVGPPEPKTTLRTNNASCIQVNL 985
Db 926 SQEESTKYNFATSDKMKKEAHRVNVNLSQDLAISINFWVPVLLNGVA-VWDVVM 984
Qy 986 EPPG-----PPVHPEELOHTNR---LNGSNTQCQVVRCHLQOLAKGTEVSGLLR 1032
Db 985 EAPQSPLCVSERKPPQHSDFLTQISRSPMLDCSIADCLQFRCDVPVSFVQEBELDTLKG 1044
Qy 1033 LVHNEPFRRAKPSLTVTSFELGTBEGSVLQLTASRWESLLE-VVOTRPILISLWIL 1091
Db 1045 NLSFGWRETLQKVLVSVVAEITFTSVYSQLPQGFQAFMRAQMEMVLEDEVTNAPII 1104
Qy 1092 IGSVLGGLLALLAVFCLMKLGRF-AHKKIPEREKREE 1128
Db 1105 MGSVGAALLLALITATLYKLGFPRHYKEMLEDKPED 1142

RESULT 9
US-08-286-889-2
Sequence 2, Application US/08286889
Patent No. 5470953
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Mich
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: P38,659
REFERENCE/DOCKET NUMBER: 27866/32168
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-286-889-2
Query Match 19.0%; Score 1126; DB 1; Length 1161;
Best Local Similarity 29.5%; Pred. No. 7 5e-99;
Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;
Qy 23 FNLDHHPLRFPQPPBAEFYSVLQHVGGQRMVLGAPWDGSGDRRDVYRCPVGAH 82
Db 17 FNLDVEEPTIFQ-EDAGGFGQSVVQF--GGSR-LVVGAPLEVVAAQGTGRLYDC---AAA 69
Qy 83 NACCAKGLHDYGLGNSSHPAVNMHGLMSLLETDDGGFMACAPLMSRACGSSVFPSSGIC 142
Db 70 TGMQCPIPL-----HIRPEAVNMSLGLTAASTNGSRLLACGPTLHRVCGENSXKSGC 123
Qy 143 ARYDASFOPOQSLAPTAQRCF-TYMDVWIVLDCSNSI--YPMSEVOTFLRLRLVGLFIDP 199
Db 124 LLGSRWEIIQTVPDATPCPHQEMDIVFLIDSGSIDQNDNFQMGKGFVQAVMGQ--FEG 181
Qy 200 EQIQVGLVQYGESPVHWSLGDPRTKBEVVVRAAKNLSRREGRETKTAQAIMVACTEFSQ 259
Db 182 TDTLFAIMQVSNLLKIHFTFTQFTSPSQSLVDPIVLQKGL-TFTATGILTVVTQLFHH 240
Qy 260 SHGCRPEAARLLVVVDGESH-DGEBLPAALKACBAGRVTTRYGIAVLGHYLRQRDPSSF 318
Db 241 KNGARKSAKKILVITDGQKYKDPLEYSQVIPAQEKAGIIRYAIGV-GHAF---QGPTA- 295
Qy 319 LRERTIASPPDERFEFFNVTDAAALTDIVDALGDRIFGLBGSNAENESSFGLBMSQIGFS 378
Db 296 RQELNTISSAPPQDQHVKNFADNFAALGSIQKQLEKIYAVEGTOSRASSSFQHEMSQGF 355
Qy 379 THRDKGILFGVMGAYDWGGSVLMLEGGHRLFPFRMALEDEFFPPALQNH-AY 430
Db 356 TALTMGDLPLGAVGSPFSW-----SGGAPLYPPNMS-----PTFINMSQENVMDR 402
Qy 431 LGYSVSMLLRGRRRLPLSGAPFRIRGKVIAP-QLKKGAVRVAOSLQEQGTSFGE 489
Db 403 LGYSTELALMKVQNLVL-GAPRYQHTKAVITQVSQW--RKKAEVTTQTGTSYFAS 459
Qy 490 LCPDTRDGTDTDLVLAAPMFLGPONKETGRVYVYLV--GQOSLLTLQGTLOPEPPQD- 546
Db 460 LCSVDVDSGSDTLILIGAPHYY--FQTRGGQVSVCPPLPRQVQWQCDVILGEQHPW 517
Qy 547 ARFGFAMGALPDNLQDGFADVAVGAPLEDHGOALYLYHG-TOSGVRPHPAQRIAAAMP 605
Db 518 GRFGAALTVLGDVNEKLDVGAIGAPQENRGAVALFHGASGSGISPSHSQRIASSQLS 577
Qy 606 HALSYFGRSVDGRDLDDGLDVAVGAQAAALLSRPIVHLTPSLVTPQASIVVQD 665
Db 578 PRLQYFQALSGGQDLTQDGLMDLAVGARGQVILLRSLPVLKVGAMRPFSEVAKAVYR 637
Qy 666 C-----RRRQEAVALCTAALCFQVTSRTPTGRWDHQFYMRFTASLDWTAGARAAFDS 718
Db 638 CWEEKPSALEAGDATVCLTIQ-----KSLDQLGDIQSSVRFDLALDPLGRLTSAI 692
Qy 719 GQRLSPRLRLSVGNVTCQQLHFHVLDTSDYLRLPVALTVTFDALDNTTKPG-----PVLN 772
Db 693 KNPTLTRKTLGLG-IHCETLKLLPDCVEDVSPPIILHLNFSLVREPIPSQNLRPVLA 751
Qy 773 EGSPTSQKLVPSKDCGPONECVTDLVQVNMMDIRSKAPFVVRGGRKVLVSTLEN 832
Db 752 VGSQDLFTASLPKCNQDGLCEGD--LGVTLSFSGLOT---LTVGSSLELNVITVMN 806
Qy 833 RKNAYNTSLSIIFSRNLHLASLTPORESP-----IKVEC-AAPSAHARL-----CSVGHPV 883
Db 807 AGDSYGVTVSVLYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLASSRCSVNHPI 866
Qy 884 FOTGAKVTFLLEPEFSCSSLLSQVFGKLTASSDSLENGTLQENTATSAIYQEPHLLF 943
Db 867 FHESNGTFTVTFDVSYSKATLGDRM-LMRASASSENKASSKATFQLELPVKYAVVTMI 925
Qy 944 S--SESTL-----HRYEVHPYGTLPVGPPEPKTTLRTNNASCIQVNL 985
Db 926 SQEESTKYNFATSDKMKKEAHRVNVNLSQDLAISINFWVPVLLNGVA-VWDVVM 984

Qy 986 EPPG-----PPVHPELOHTNR---LNGSNTQCVVRCHLQGLAKGTVEVSGLLR 1032
Db 985 EAPQSILPCVSRKPPQHSDFLTQISRSPMLDCSIADCLQPCDVPVSQVEELDTLKG 1044
Qy 1033 LVHNEPFRRAKFKSLTVVSTFELGTSGSVLQLTEASRWSLSLE-VVQTRPILISLWIL 1091
Db 1045 NLSFGWRETLQKVLVSVVAEITFTDSVSVQLPGQAFMRAQMEMVLEDEVDVNAIPII 1104
Qy 1092 IGSVLGGLLLALLVFLWLKLGFF-AHKKIPEEBKREE 1128
Db 1105 MGSSVGALLLALLATATLYKLKLGFFKRYKEMLEDKPED 1142

RESULT 10

US-08-485-618-2
Sequence 2, Application US/08485618
Patent No. 5728533
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-2

Query Match 19.0%; Score 1126; DB 1; Length 1161;
Best Local Similarity 29.5%; Pred. No. 7.5e-99;
Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;

Qy 23 FNLDHHPRLFPPEAEFGYSVLQHVGGQGMVLPVGPWDPGSDRRGDVYRCPVGAH 82
Db 17 FNLDVEETIFQ-EDAGCGFGSVQVF--GGR-LVVGAPLEVAANQTRLYDC----AAA 69
Qy 83 NAPCAKHLGDIYQLGNSHPVNMHLGMSLLETDDGGFMACAPLWSRACGSSVFSSGIC 142

Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASTNGSRLLACGPTLHRVCGENSYSKGSC 123
Qy 143 ARVDASFPQPOCSLAPTAQRCP-TYMDVVIVLDGSNSI--YPMSEVQOTFLRLRVGLKFLDP 199
Db 124 LLLGSRWEIIQTVPDAPTECPHQEMDIVFLIDGSGSIDQDNQMGKFGVQAVMQ--PEG 181
Qy 200 EQQVGLVQVGEFVHWSLGDFTKEEVVRAAKNLSRREGRETAKTAQALMVACTEFGSQ 259
Db 182 TDTLFAIMOYSNLLKIHFTFTQRTSPSQSLVDPIVOLKGL-TFTATGILTVVTQLPFH 240
Qy 260 SHGGRPEAARLVVVVTDGESH-DGEELPAALAKACEACRVTTRYGIAVLGHVLRRLRRDSSP 318
Db 241 KNGARKSAKKILVITDQKTKDPLEYSDVIPQEKAGIIRYALGV-GHAP---QGPTA- 295
Qy 319 LREIRTIASDPDERFFNFVVTDEAALTDIVDALDGFLEGSCHAENSSSGFLEMSQIGFS 378
Db 296 RQELNTISSAPPQDHVFKVDNFAALGSIQKLOQEKIYAVEGTQSRASSSFQHEMSQEGFS 355
Qy 379 THRLKDGILFGMVGAYDNGSVLWLEGCHRLFPPRMALEDEFPFALQNH-----AY 430
Db 356 TALTMDFLGLGAVGFSW-----SGGAFLYPPNMS-----PTFINMSQENVDMRDSY 402
Qy 431 LGYSVSSMLRGGRRLFLSGAPRFRHGKVIAR-OLAKKDGAVRVAQSLQSGQISGYFSGE 489
Db 403 LGYSTELALWKGVQNLVL-GAPRYQHTGKAVIFTVQSRQW--RKAETVGTQISGYFGAS 459
Qy 490 LCPDLTDRTDGTVDLLVAAPMFLGPQNKETGRVVVYLV--GQSSLLTLQGTLOPEPPD- 546
Db 460 LCSVDVDSGSDTLILIGAPHYY--EQTRGQVSVCPPLRGQRVQWQCDVALRGEQHPW 517
Qy 547 ARFGFANGALPDNLQDGFADVAGCAPLEDHQGALYLYHG-TQSGVRPHPAQRTAAASMP 605
Db 518 GRFGAALTVLGDVNEKLIIDVAIGAPGEQENRGAVILFHGASESGISPSHQRIASSQLS 577
Qy 606 HALSYFGRSVDGRDLDDGLVDVAVGAQGAAILLSRPIVHLTPSLVETPQALSVVQRD 665
Db 578 PRLOYFGQALSGGQDLTQDGLMDLAVGARGQVLLRLSLPLVKGVAMRFSPEVAKAVYR 637
Qy 666 C-----RRRGQEAVALCLTAALCFQVTSRTPTGRWDHOFYMRFTASLDEWTAAGAAFDGS 718
Db 638 CWEEKPSALEAGDATVCLTIQ-----KSLDQLDGIQSSVRFDLALDGRILTSRAIFNET 692
Qy 719 GQRLSPRLRLSVGNVTCQQLHFHVL-DTSLYLPVALVTTFALDNTTKPG-----PVLN 772
Db 693 KNPTLTRRKTGLG-IHCETLKLALLPCDVEDVWSPFIILHNFSLVREP IPSPQNLRPVLA 751
Qy 773 EGSPTSIQKLVPSKDCGPNCEVTDVLQVNMIDIRGSRKAPFVVRGRRKVLVSTTLEN 832
Db 752 VGSQDLFTASLPFPEKNCQDGLCEGD--LGVTLSFSGLT---LTVGSSLSLNVIVTVNM 806
Qy 833 RKENAYNTSLSIIPSRNLHLASLTPQRESP-----IKVEC-AAPSAHARL-----CSVGHPV 883
Db 807 AGEDSYGTVVSLLYYPAGLSHRRVSGAQKHQSALRLACETVPTDEGLSRSSCVNHP 866
Qy 884 FQTGAKVTELLPEFSCSSLLSQVFGKLTASDSLERNGLTQENTQATSAVIOEPHLLF 943
Db 867 FHEGNGTFIVTFDVSFKATLGDRLM-LMRASASSENKASSKATFQLELPVKVAVYTM 925
Qy 944 S--SESTL-----HRYEVHPYGTLPVGPPEEFTTLRTNNAFCIVQNL 985
Db 926 SRQESTKYFNFATSDKKMKEAHRVNVNLSORDLAISINFWVPVLLNGVA-VMDVVM 984
Qy 986 EPPG-----PPVHPELOHTNR---LNGSNTQCVVRCHLQGLAKGTVEVSGLLR 1032
Db 985 EAPQSILPCVSRKPPQHSDFLTQISRSPMLDCSIADCLQPCDVPVSQVEELDTLKG 1044
Qy 1033 LVHNEPFRRAKFKSLTVVSTFELGTSGSVLQLTEASRWSLSLE-VVQTRPILISLWIL 1091
Db 1045 NLSFGWRETLQKVLVSVVAEITFTDSVSVQLPGQAFMRAQMEMVLEDEVDVNAIPII 1104
Qy 1092 IGSVLGGLLLALLVFLWLKLGFF-AHKKIPEEBKREE 1128
Db 1105 MGSSVGALLLALLATATLYKLKLGFFKRYKEMLEDKPED 1142

Db 1105 MGSSVGALLLALITATLYKLGFFKRYKEMLEDKPED 1142

RESULT 11

US-08-652-2

; Sequence 2, Application US/08362652

; Patent No. 5766850

; GENERAL INFORMATION:

; APPLICANT: Gallatin, W. Michael

; APPLICANT: Van der Vieren, Monica

; TITLE OF INVENTION: No. 576850el Human 2 Integrin Alpha Subunit

; NUMBER OF SEQUENCES: 93

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 233 South Wacker Drive, 6300 Sear Tower

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States

; ZIP: 60606-6402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/362,652

; FILING DATE:

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/173,497

; FILING DATE: 23-DEC-1993

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/286,889

; FILING DATE: 5-AUG-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Williams Jr., Joseph A.

; REGISTRATION NUMBER: 38,659

; REFERENCE/DOCKET NUMBER: 27866/32391

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-474-6300

; TELEFAX: 312-474-0448

; TELE: 25-3856

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1161 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-362-652-2

Query Match 19, 0%; Score 1126; DB 1; Length 1161;

Best Local Similarity 29.5%; Pred. No. 7.5e-99;

Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;

Qy 23 FNLDHPRFPFGPEAEFGYSLQHVGGQRMVLGAPMDPGSDRRGDVYRCPVGAH 82

Db 17 FNLDVEPTIFQ-EDAGFGSVVQF--GGSR-LVVGAPLEVVAANTQGRLYDC--AAA 69

Qy 83 NAPCAKHLGDYQGNSSHPANVHLGMSLLETGDDGFWACAPLWSRACSSVFFSGIC 142

Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTAASTNGSRLLACGPTLHRVCGENSYSKGC 123

Qy 143 ARVDASFPQGSGLAPTAQRCB-TYMDVVIVLDGNSI--YPMSEVQTFRLRLVCKLFDIP 199

Db 124 LLLGSRWEIIQTVPDAPTECPHQSDMDIVFLIDGSGSIDQDNFQMKGFVQVMQ--FEG 181

Qy 200 EQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRETAKQAINVACTEGFSQ 259

Db 182 TDTLFPALMQYNSLKKIHFTFTQRTSPSQSLVDPIVLQKGL-TFTATGILTVVTQLFHH 240

Qy 260 SHGGRPEARLLVVVTDGESH-DEEELPAALAKACEAGRVTRYGTAVLGHYLRDRDPSF 318

Db 241 KNGARKSAKKILVITDQKYKDPLEYSVDVIPQAEKAGIIRYAGV-GHAF--QGPTA- 295

Qy 319 LREIRTIASDPDERFFFNVTDEAALFDIVDALGDRIFGLEGSHAENESSFGLMSIQGFS 378

Db 296 RQELNTISSAPPQDHFVKVDNFAALGSIQKQLEKIYAVEGTQSRASSSFQHEMSQEGFS 355

Qy 379 THRLKDGILFGMAYDWMGGSVLMLEGGHRLFPFPRMALEDEFPFPAQNH-AY 430

Db 356 TALTMDELFLGAVGFSW-----SGGAPLYPPNMS-----PTFINMSQENVMDRDSY 402

Qy 431 LGYSVSSMLLRGRRRLFLSGAPRHRGKVIAF-QLKDGAVRVAQSLQEQIGSYFGSE 489

Db 403 LGYSTELALWKGVQNLVL-GAPRIQHTGKAVITQVSROW--RKKAEVGTQIGSIFGAS 459

Qy 490 LCPLDTRDGTDTVLVLAAPMFLGPNQKQKGTGRVYVYL--GQOSLLTLQOTLPEPPQD- 546

Db 460 LCSVDVDSGSDTLILIGAPHYY--BQTRGGQSVSCPLPRGQVQWQCDVLRGEQHPW 517

Qy 547 ARFGFANGALPDNLQDGFADVAVGAPLEDHGHQALYLYHG-TOSGVRPHAPQRIAAAMP 605

Db 518 GRFGAALTVLGDVNEKLDLIDVAIGAPGEQENRGAVYLFHGASRSGISPSHSQRIASSQLS 577

Qy 606 HALSYFGSRVDGRDLDDGDLVDVAVGAOGAATLLSSRPVHLTPSLVTPQAISVVQRD 665

Db 578 PRQYFGQALSSGGQDLTQDGLMDLAVGARGQVLLSLPLVKVGVAMRFSFVEVAKAVYR 637

Qy 666 C-----RRRGQEAVALTAALCFQVTSRTPGSRWDHQFYMRFTASLDDEWTAGARAAFDGS 718

Db 638 CWEEKPSALEAGDATVCLTIQ-----KSSLDQLDGIQSSVRFDLALDPGLTSRAIFNET 692

Qy 719 GQBLSPRLRLSVGNVTCBQLHFFHVD-TSDYLRPVALVTFPDLNNTKFG-----PVLN 772

Db 693 KNPTRLRRKTLGLG-IHCETLKLLPDCVEDVVSPIILHLNLSLVRPIPSPONLRPVLA 751

Qy 773 EGSPTSIOKLVPFSKDCGPDNECVTLVLQVNMDIRSKAPFVVGGRKKVLVSTTLEN 832

Db 752 VGSQDLFTASLPFEKNCQDGLCEGD--LGVTLFSGLQT---LTVGSSLELNVIVTVN 806

Qy 833 RKENAYNTSLIIFSRNLHLASLTPORES-----IKVEC-AAPSAHARL---CSVGHVP 883

Db 807 AGEDSYGTVVSLLYPAGLSHRRVSGAQKQPHQSALACETVTEDEGLSSRCSVNHP 866

Qy 884 FQTGAKVTFLLRPFECSSLLSQVFGKLTASSDSLRNGTLQENTACTSAYIQEYHLLF 943

Db 867 FHGSGNGTFIVTFDVSFKATLGDRM-LMRASSENKASSKATFQLELPVYAVYTM 925

Qy 944 S--SESTL-----HRYEVHPYCTLVPGCPPEKTTLTNTNASCIVQNL 985

Db 926 SROEESTKYPNFATSDKMKAEHRYRVNLSQDLAISINFWVPVLLNGVA-VMDVVM 984

Qy 986 EPPG-----PPVHPELQHTNR---LNGSNTQCQVVRCHLQLAGTEVSUGLLR 1032

Db 985 EAPSQSLPCVSEKPKQHSDFLTQISRSPMLDCIADCLQFCDCVPSFVQEBLDTLKG 1044

Qy 1033 LVNNEFPRAKFKSLTVVSTFELGTBEGSVLQTEASRWSESLLE-VVQTRPLISLWIL 1091

Db 1045 NLSFGMWRETQLQKVLVSVAEITFDTSVYSQLPQGEAFMRAQMEMVLEEDVYNAIP 1104

Qy 1092 IGSVLGILLALLVFCMLKLGFF-AHKKIPSEKREE 1128

Db 1105 MGSSVGALLLALITATLYKLGFFKRYKEMLEDKPED 1142

RESULT 12

US-08-605-672-2

; Sequence 2, Application US/08605672

; Patent No. 5817515

; GENERAL INFORMATION:

; APPLICANT: Gallatin, W. Michael

; APPLICANT: Van der Vieren, Monica

; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit

; NUMBER OF SEQUENCES: 103

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 48.3117 Seconds
(without alignments)
1749.117 Million cell updates/sec

Title: US-09-647-544-4
Perfect score: 5932
Sequence: 1 MELPFVTHLFLPLVFLTGLC.....GFFAHKKIPBEKREKLEQ 1132

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES *

Result No.	Score	Query Match	Length	ID	Description
1	2345	39.5	1217	4	US-09-949-016-7892
2	1884	31.8	1180	4	US-09-000-004A-2
3	1775.5	29.9	1183	4	US-09-532-310B-5
4	1733.5	29.2	1181	4	US-09-000-004A-4
5	1733.5	29.2	1181	4	US-09-949-016-6189
6	1733.5	29.2	1195	4	US-09-949-016-10747
7	1704.5	28.7	1183	4	US-09-532-310B-6
8	1126	19.0	1161	1	US-08-173-487-2
9	1126	19.0	1161	1	US-08-286-889-2
10	1126	19.0	1161	1	US-08-485-618-2
11	1126	19.0	1161	1	US-08-362-652-2
12	1126	19.0	1161	2	US-08-605-672-2
13	1126	19.0	1161	2	US-08-482-293A-2
14	1126	19.0	1161	2	US-08-943-363-2
15	1126	19.0	1161	3	US-09-193-043-2
16	1126	19.0	1161	4	US-09-688-307A-2
17	1126	19.0	1161	4	US-09-350-259-2
18	1124	18.9	1161	3	US-09-193-043-55
19	1124	18.9	1161	4	US-09-688-307A-55
20	1124	18.9	1161	4	US-09-350-259-55
21	1120.5	18.9	1161	1	US-08-485-618-99
22	1120.5	18.9	1161	2	US-08-605-672-99
23	1120.5	18.9	1161	2	US-08-482-293A-99
24	1120.5	18.9	1161	2	US-08-943-363-99
25	1120.5	18.9	1161	3	US-09-193-043-99
26	1120.5	18.9	1161	4	US-09-688-307A-99
27	1120.5	18.9	1161	4	US-09-350-259-99

28	1118	18.8	1161	1	US-08-485-618-55	Sequence 55, Appl
29	1118	18.8	1161	1	US-08-362-652-55	Sequence 55, Appl
30	1118	18.8	1161	2	US-08-605-672-55	Sequence 55, Appl
31	1118	18.8	1161	2	US-08-482-293A-55	Sequence 55, Appl
32	1118	18.8	1161	2	US-08-943-363-55	Sequence 2, Appl
33	1115.5	18.8	1170	2	US-08-789-078-2	Sequence 2, Appl
34	1115.5	18.8	1170	2	US-08-752-633-2	Sequence 2, Appl
35	1115.5	18.8	1170	5	PCT-US95-04886-2	Sequence 2, Appl
36	1114.5	18.8	1161	1	US-08-485-618-53	Sequence 53, Appl
37	1114.5	18.8	1161	1	US-08-362-652-53	Sequence 53, Appl
38	1114.5	18.8	1161	2	US-08-605-672-53	Sequence 53, Appl
39	1114.5	18.8	1161	2	US-08-482-293A-53	Sequence 53, Appl
40	1114.5	18.8	1161	2	US-08-943-363-53	Sequence 53, Appl
41	1114.5	18.8	1161	3	US-09-193-043-53	Sequence 53, Appl
42	1114.5	18.8	1161	4	US-09-688-307A-53	Sequence 53, Appl
43	1114.5	18.8	1161	4	US-09-350-259-53	Sequence 53, Appl
44	1108.5	18.7	1155	1	US-08-286-889-46	Sequence 46, Appl
45	1108.5	18.7	1155	1	US-08-485-618-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1

US-09-949-016-7892
; Sequence 7892, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7892
; LENGTH: 1217
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-7892

Query Match		39.5%	Score 2345;	DB 4;	Length 1217;
Best Local Similarity		42.2%	Pred. No. 1.3e-216;		
Matches		505;	Conservative 204;	Mismatches 407;	Indels 80; Gaps 16;
Qy	1	MELPFVTHLFLPLVFLTGLCSPFNLDHHPRLFPGPPEAEFGSVLQHVGGGQRMVLVGA	60		
Db	30	MDLPRGLVVAWALSILWPGFTDTFNMDTKPRVIFGSRTAFFGYTVQOHDISGNKWLVLVGA	89		
Qy	61	PWDGPGDRDRYVRCVPVGGAHNAPCAKHGLGDYQLGNSSHPANMHLGMSLLETDDGGG	120		
Db	90	PLETNGYQKTGDVYKCPV---IHGNCCTKLNLRVTLNSVSRKDNMRGLSLATNPKDNS	146		
Qy	121	FMACAPLWSRACGSSVFSSGICARVDASFPQGSIAPTAQCPTVMVIVLDGNSIYP	180		
Db	147	FLACSPLSWHECGSSYTTGMCNRWNGNFRFSKTVAPALQRCQTMIVIVLDGNSIYP	206		
Qy	181	WSEVQTLRRLVGLKFLFDPEQIQVGLVQYQSPVHWSLGDGFRTEKVEVVRRAKNLSRREG	240		
Db	207	WVEVQHFLINILKFFYIGPGQIQGVVQYQGVEDVHFEHLNDYRSVKDVVEASHLEQRCG	266		
Qy	241	RETKTAQIMVACTGFSQSHGSRPEARLAVVVTVDGSHDGEELPAALKACEAGRVTY	300		
Db	267	TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGSHDSDPLEKVIQQRDNVTRY	324		
Qy	301	GIADVILYLRQRDPSSFLREIRTIASDPDRFPFNVTDAAALTDIVDALGDRIFGLSGS	360		

Db 783 PENCPLDLPNSVHEHPPAKCGNKKERCISDLTLNVST-----TEKSLIIVKSQHDKF 838
Qy 825 LVSTTLNKRKENAYNTSLSIIFSRNLHLATP-QRESPIKVECAAPSAHARLCSVGHVP 883
Db 839 NVSLTVNKNKGSAYNTRTVQHSPLNIFSGIEEIQDS-----CESNQNTCRVGVYP 891
Qy 884 FOTGAKVTFLLEPEFSCSLLSQVFGKLTASSDSLRNGTLQENTTAQTSAYIQEPHLLF 943
Db 892 LRAGETVTFKIIFOFNTSHLSENAIHLSATSDSEEPLESNDNEVNISIPVKYEVGLQF 951
Qy 944 SSESTLHRYEHPYGTLP-----VG-----PGPEFKTL----- 972
Db 952 YSSASEHHISVAANETIPEFINSTEDIGNEINVPYTIKRGHPMPPELQLSIFPNLTAD 1011
Qy 973 -----RTNNASCIQNLTPEPG-----PPVHPEELOHTNRLNGSNTQCQVRC 1015
Db 1012 GYPVLYPIGWSSSDNVNCRPSLEDPPFGINSKKMTISKSEVLKRGTIQDCSSTCGVATI 1071
Qy 1016 HLGOLAKG-TEVSVGLRLVHNEPFRRAKFKSLTVSTFELGTGEGSVLQLTEASRWSES 1074
Db 1072 TCSLLPSDLSQVNSL--LWKPTFIRAHFSSNLTLRGELKSENS--LTLSSNRKREL 1128
Qy 1075 LLEVQOT-RPTLISILWILGVLGGLLALLVFLCKLGFPAHKKIPPEEKREE 1128
Db 1129 AIQSKDGLFORVPLWILLSAPAGLLMLLILALWKIGF---KRPLKKMEK 1180

RESULT 3

US-09-532-310B-5
Sequence 5, Application US/09532310B
Patent No. 6596276

GENERAL INFORMATION:

APPLICANT: Senger, Donald R
Detmar, Michael
Claffey, Kevin P

TITLE OF INVENTION: Method for inhibiting tumor
angiogenesis in a living subject

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: David Prashker, Esq.

STREET: P.O. Box 5387

CITY: Magnolia

STATE: Massachusetts

COUNTRY: USA

ZIP: 01930

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage

COMPUTER: Dell PC

OPERATING SYSTEM: MS DOS

SOFTWARE: Microsoft Word version 97

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/532,310B

FILING DATE: 22-Mar-2000

CLASSIFICATION: Unknown

ATTORNEY/AGENT INFORMATION:

NAME: David Prashker, Esq.

REGISTRATION NUMBER: 29,693

REFERENCE/DOCKET NUMBER: BIS-036

TELECOMMUNICATION INFORMATION:

TELEPHONE: (978) 525-3794

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1183 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-532-310B-5

Query Match

Best Local Similarity 29.9%; Score 1775.5; DB 4; Length 1183;

34.8%; Pred. No. 1.4e-161;

Matches 416; Conservative 219; Mismatches 449; Indels 111; Gaps 24;
Qy 13 LVFLTGLCSFNLDEHHRLPFGPPEAEFGYSVLQHVGGGORMMLVAGAPWDGSGDRGD 72
Db 21 LTVILGFCVSNVDVKMSMTFSGPVEDMFGYTVQOYENEEGKWVLIGSPLVGQPKNRGT 80
Qy 73 VYRCPVGAHNAPECAKHLGDYQLGNSSHPAV-----NMHLGMSLLTDDGGGFMACAPL 127
Db 81 VYKCPVGRGSLPCVKLDLP-----VNTSIPNVTVEKENMTFG-STLVNPNNGGLACGPL 135
Qy 128 WSRACGSVSSSGTCARVDASFQPGSLAPTAQRCPTYMDVIVLDGNSNIYPMSEVQTF 187
Db 136 YAYRCGHLHTTGICSDVSPTFQVNSIAP-VQESCXTKLDIVIVLDGNSNIYPMDSV-TA 193
Qy 188 LRLVLKGLFTDEQIQVLQVYGESPVHWSLGDFTKKEEVRRAAKLSRREGRETAKQ 247
Db 194 LNDLLKEMDLPKXTXGVIXYGENVTHFNLNKYSSTEELVAAKKIVXRGGRXTMTAL 253
Qy 248 AIMVACTEGFSQSHGGRPEARLLVVVTDGESHDEELPAALKACEAGRVTRYGI AVLGH 307
Db 254 GTDTARKEATEARGARGVKVMVITDGEHDXNHLKKVIGDCEDENIXRFSAIILGS 313
Qy 308 YLRRQDPSSFLREIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGSHAENESS 367
Db 314 YNRGNLSTEFVEBEIKSIASEPTKSFNVSDDELALVTIVKTGLGERIFALEATADXSAAS 373
Qy 368 FGLEMSOIGSTHRLKDGILFGMVGAYDWGGSVLMLEGGHRLFPPRMALEDEFPPALQNH 427
Db 374 FEMEMSGTGFSAHYSQDMVWLGA VGADWNGTVVYNQASQIIIPRNTTFNVESTKQNEPL 433
Qy 428 AAVLGYSVSMLLRGGRRLFLSGAPRFRHKGKVIATFOLKXKDGAVRVAQSGOGEIGSYFG 487
Db 434 ASYLGYTVNSATASSGDVLYIAGQPRYNHTGQVYIYRM-BEGNIKIQTLSGXIGSYFG 492
Qy 488 SELCPLDTRDGTDLVLLVAAPMPLGPONKETGVYVYLVGQQSLLTLTQGTLOP-EP--- 543
Db 493 SILTTTIDIKDSNTNILLVGAPWYMGTEKEGGKVYVYAL-NQRFYQVMSLAPMEPIKQ 551
Qy 544 -----PQDARPGFAMGALPDNLQDGFADVAVGAPLEDHGOGALYLYH 585
Db 552 TCCSSRQHNSCTTENKNEPCGARFGTAIAAVKOLNLDGNDIVIGAPLEDHGGAVYIYH 611
Qy 586 GTQSGVRPHPAQRITAAASMPHALSYFGRSVDGRLLDGDLDLVAVAGQAQGAAILSSRP 645
Db 612 GSGKTIKRYAXRIPSGDGKTLKFQGSITHEMDLNGDGLTDVTICGLGAALFWSRDV 671
Qy 646 VHLTPSLEVTPOATISVVQDRCRRRGOEAVCLTAALCFQVTSRTPGRDWHQFMRFTASLD 705
Db 672 AVVKVTWNPENKVNIOKKNHMEGKETVCINATVCEVFKLSKEDITYEADLQYRVTL 731
Qy 706 EWTAGARAAFDGGQRLSPRLRLSVGNVTCEQLHFLHVLDTSDYLRLPVALTFTFALDNTT 765
Db 732 SLRQISRSFSGTQERKVQR-NITVRKSECTKSHFYMLDKHPQDSVRITLDF---NLT 786
Qy 766 KP--GPVLNEGSPTSIOKLVPFSSKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGGRK 823
Db 787 DPENGVLDDSLPNSVHEIYTPFAKDCGNKCKISDLSLHV-----ATTEKDLILVRSQNDK 842
Qy 824 VLVSTTLLENKENAYNTSLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVP 883
Db 843 FNVSLTVNKTDSAYNTRTVHYSPLNVFSGI-----EAIQKDC--FSNHNITCKVGVYP 896
Qy 884 FOTGAKVTFLLEPEFSCSLLSQVFGKLTASSDSLRNGTLQENTTAQTSAYIQEPHLLF 943
Db 897 LRRGEMVTFKILFOFNTSYLMENVTVILSATSDSEEPETLSDDNVNVISIPVKYEVGLQF 956
Qy 944 SSESTLHRYEHPYGTLP-----VG-----PGPEFKTL----- 972
Db 957 YSSASEYHISIAANETVPEVINSTEDIGNEINIFYLIRKSGSFPMPPELKLISFPNMTSN 1016
Qy 973 -----RTNNASCIQNLTPE-----PPVHPEELOHTNRLNGSNTQCQVRC 1015
Db 1017 GYPVLYPTGLSSSENANCRPHIFEDPFSINSKGKQNTTSTDHLKRGTLDCNCTKFAITC 1076

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 0.766641 Seconds
(without alignments)
2761.097 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFFAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: Piri:.*
2: Piri2:.*
3: Piri3:.*
4: Piri4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	46.5	214	2 A11073	probable membrane
2	51	44.7	672	2 G69503	signal-transducing
3	51	44.7	856	2 H64552	endopeptidase Clp
4	51	44.7	1045	2 S60571	integrin alpha v c
5	50	43.9	357	2 AC1104	B. subtilis yacI p
6	50	43.9	505	1 S77034	protein kinase pkn
7	50	43.9	1034	2 A36108	integrin alpha-v c
8	50	43.9	1044	2 T10050	integrin alpha-5 c
9	50	43.9	1048	2 A27421	probable lipoprote
10	49.5	43.4	236	2 T12766	outer capsid prote
11	49.5	43.4	959	1 B60017	B. subtilis yacI p
12	49	43.0	357	2 AC1466	hypothetical prote
13	49	43.0	615	2 T06108	hypothetical prote
14	49	43.0	853	2 S74279	probable endopepti
15	49	43.0	856	2 C71956	M2 protein precurs
16	48.5	42.5	407	2 S23325	hypothetical prote
17	48	42.1	421	2 H90433	squalene monooxyge
18	48	42.1	573	2 A55767	G protein-coupled
19	48	42.1	962	2 JC5808	bazooka gene prote
20	48	42.1	1464	2 T13716	DNA topoisomerase
21	48	42.1	1526	2 JN0598	hypothetical prote
22	47	41.2	102	2 E75077	flagellar switch p
23	47	41.2	112	2 E70433	hypothetical prote
24	47	41.2	206	2 T16946	hypothetical prote
25	47	41.2	512	2 AC3203	IS3 family transpo
26	47	41.2	512	2 AD2835	IS3 family transpo
27	47	41.2	512	2 AD3049	probable transposa
28	47	41.2	512	2 G97622	probable transposa
29	47	41.2	512	2 G98236	probable transposa

30	47	41.2	692	2 S37976	hypothetical prote
31	47	41.2	861	2 T02267	trehalose-6-phosph
32	46	40.4	278	2 D84492	hypothetical prote
33	46	40.4	308	2 S11153	oligopeptide trans
34	46	40.4	308	2 E95220	hypothetical prote
35	46	40.4	323	2 E98084	hypothetical prote
36	46	40.4	450	2 S37900	hypothetical prote
37	46	40.4	544	2 AH2971	hypothetical prote
38	46	40.4	544	2 E98311	probable ATP-bindi
39	46	40.4	847	2 C96703	hypothetical prote
40	46	40.4	1037	2 A60163	glycoprotein Iib -
41	46	40.4	1407	1 S28589	trichohyalin - rab
42	45.5	39.9	570	2 F70332	proline-tRNA synth
43	45	39.5	273	2 T47612	hypothetical prote
44	45	39.5	325	2 T18283	hypothetical prote
45	45	39.5	333	2 T08850	alternative respir

ALIGNMENTS

RESULT 1

A11073
probable membrane protein smp [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: A11073
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: A11073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-214 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD03408.1; PID:gl6505677; GSPDB:GN00176
C:Genetics:
A:Gene: smp

Query Match 46.5%; Score 53; DB 2; Length 214;
Best Local Similarity 71.4%; Pred. No. 5.6;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 6 AHKKIPEEKREK 19
Db 200 ASKPVPEEREK 213

RESULT 2

G69503
signal-transducing histidine kinase homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: G69503
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Nature 390, 364-370, 1997
A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69503; MUID:98049343; PMID:9389475
A:Accession: G69503
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-672 <KLE>
A:Cross-references: UNIPROT:O28247; GB:AE000963; GB:AE000782; NID:g2689286; PIDN:AAB8922

Query Match 44.7%; Score 51; DB 2; Length 672;

Best Local Similarity 55.6%; Pred. No. 33;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 FAHKKIPPEEKREKLEQ 22
|||:|||||:|
Db 572 FAFKMEDERRRELLKQ 589
|||:|||||:|

RESULT 3
H64552
endopeptidase Clp ATP-binding chain B - Helicobacter pylori (strain 26695)
N/Alternate names: ATP-dependent Clp proteinase regulatory chain
N/Contents: adenosinetriphosphatase (EC 3.6.1.3)
C/Species: Helicobacter pylori
C/Date: 09-Aug-1997 #sequence_revision 15-Aug-1997 #text_change 09-Jul-2004
C/Accession: H64552
R/Tomb, J.F.; White, O.; Karlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A/Reference number: A64520; MUID: 97394467; PMID: 9252185
A/Accession: H64552
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-856 <TOM>
A/Cross-references: UNIPROT:P71404; GB:AE000545; GB:AE000511; NID:g2313349; PIDN:AAD0733
C/Function:
A/Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperone activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller peptides
C/Superfamily: endopeptidase Clp ATP-binding chain
C/Keywords: ATP; duplication; hydrolase; molecular chaperone; nucleotide binding; P-loop
F/204-211/Region: nucleotide-binding motif A (P-loop)
F/271-276/Region: nucleotide-binding motif B
F/605-612/Region: nucleotide-binding motif A (P-loop)
F/673-678/Region: nucleotide-binding motif B
F/210/Binding site: ATP (Lys) #status predicted
F/611/Binding site: ATP (Lys) #status predicted

QY 3 GFFAHKKIPPEEKREKLEQ 22
|||:|||||:|
Db 496 GEIYSKIPENKKKEELQR 515
|||:|||||:|

RESULT 4
S60571
integrin alpha v chain precursor - Iberian ribbed newt
C/Species: Pleurodeles waltliffi (Iberian ribbed newt)
C/Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C/Accession: S60571
R/Alifanari, D.; Whittaker, C.A.; Desimone, D.W.; Darribere, T.
Dev. Biol. 170, 249-261, 1995
A/Title: Integrin alpha-(v) subunit is expressed on mesodermal cell surfaces during amphibian gastrulation
A/Reference number: S60571; MUID: 95377519; PMID: 7649360
A/Accession: S60571
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-1045 <ALF>
A/Cross-references: UNIPROT:Q91292; EMBL:X81108; NID:g1008137; PIDN:CAA57014.1; PID:g1008137
C/Superfamily: integrin alpha-2b chain
F/1-29/Domain: signal sequence #status predicted <SIG>
F/30-1045/Product: integrin alpha v chain #status predicted <MAT>

Query Match 44.7%; Score 51; DB 2; Length 1045;
Best Local Similarity 42.9%; Pred. No. 51;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPPEEKREKLE 21
|||:|||||:|

Db 1014 KFGFFKVRPPQEQREQLQ 1034
|||:|||||:|

RESULT 5
AC1104
B. subtilis YacL protein homolog lmo0234 [imported] - Listeria monocytogenes (strain EGD-80)
C/Species: Listeria monocytogenes
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AC1104
R/Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Barche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussutget, O.; Entian, K.D.; Feihl, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maicournam, A.; Matok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A/Reference number: AB1077; MUID: 21537279; PMID: 11679669
A/Accession: AC1104
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-357 <GLA>
A/Cross-references: UNIPROT:Q48762; GB:NC_003210; PIDN:CAD00761.1; PID:g16409599; GSPDB:16409599
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: lmo0234
C/Superfamily: conserved hypothetical protein yacL

Query Match 43.9%; Score 50; DB 2; Length 357;
Best Local Similarity 71.4%; Pred. No. 24;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 AHKKIPPEEKREK 19
|||:|||||:|
Db 146 AKRTPPEEKTEK 159
|||:|||||:|

RESULT 6
S77034
protein kinase pKna (EC 2.7.1.-), 55K - Synecocystis sp. (strain PCC 6803)
N/Alternate names: protein ell10776
C/Species: Synecocystis sp.
A/Variety: PCC 6803
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: S77034
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shampo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, S.
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp.
A/Reference number: S74322; MUID: 97061201; PMID: 8905231
A/Accession: S77034
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-505 <KAN>
A/Cross-references: UNIPROT:P54735; EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BAAL0779
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C/Genetics:
A/Gene: pKna
C/Superfamily: Synecocystis sp. protein kinase pKna, 55K; protein kinase homology
C/Keywords: phosphotransferase; protein kinase
F/7-268/Domain: protein kinase homology <KIN>

Query Match 43.9%; Score 50; DB 1; Length 505;
Best Local Similarity 58.8%; Pred. No. 34;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKIPPEEKREKLEQ 22
|||:|||||:|
Db 376 AEQKIAENKQRELEQ 392
|||:|||||:|

RESULT 7

A:Reference number: S47541; MUID:94368864; PMID:7522056
A:Accession: S47541
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-61 <DON>
A:Cross-references: EMBL:U07375
R:Fitzgerald, L.A.; Poncz, M.; Steiner, B.; Rall Jr., S.C.; Bennett, J.S.; Phillips, D.R.
Biochemistry 26, 8158-8165, 1987
A:Title: Comparison of cDNA-derived protein sequences of the human fibronectin and vitronectin
A:Reference number: A90526; MUID:88163472; PMID:2450560
A:Accession: B29418
A:Molecule type: mRNA
A:Residues: 1-433 <FIT>
R:Suzuki, S.; Argraves, W.S.; Pytela, R.; Arai, H.; Krusius, T.; Pierschbacher, M.D.; Rukeyser, M.
Proc. Natl. Acad. Sci. U.S.A. 83, 8614-8618, 1986
A:Title: cDNA and amino acid sequences of the cell adhesion protein receptor recognizing vitronectin
A:Reference number: A26482; MUID:87041504; PMID:2430295
A:Accession: A26482
A:Molecule type: mRNA
A:Residues: 413-1048 <SU2>
R:Cheresh, D.A.; Smith, J.W.; Cooper, H.M.; Quaranta, V.
Cell 57, 59-69, 1989
A:Title: A novel vitronectin receptor integrin (alpha-v beta-x) is responsible for distribution of alpha-v beta-x
A:Reference number: A32287; MUID:89195223; PMID:2467745
A:Accession: A32287
A>Status: preliminary
A:Molecule type: protein
A:Residues: 31-35,'X',37-41 <CHE>
A:Experimental source: melanoma cell M21
A:Accession: B32287
A>Status: preliminary
A:Molecule type: protein
A:Residues: 31-33,'X',35,'X',37-41 <CH2>
A:Experimental source: lung carcinoma cell UCLA-P3
R:Lam, S.C.T.; Plow, E.F.; D'Souza, S.E.; Cheresh, D.A.; Frelinger III, A.L.; Ginsberg, M.H.
J. Biol. Chem. 264, 3742-3749, 1989
A:Title: Isolation and characterization of a platelet membrane protein related to the vitronectin receptor
A:Reference number: A30298; MUID:89139425; PMID:2465293
A:Accession: A30298
A:Molecule type: protein
A:Residues: 31-35,'X',37-40 <LAM>
R:Smith, J.W.; Cheresh, D.A.
J. Biol. Chem. 265, 2168-2172, 1990
A:Title: Integrin (alpha-vbeta-3)-ligand interaction. Identification of a heterodimeric alpha-vbeta-3 ligand
A:Reference number: A35035; MUID:90130470; PMID:1688848
A:Accession: A35035
A>Status: preliminary
A:Molecule type: protein
A:Residues: 66-72;169-171,'X',173-176;221-230;255-258,'X',260;325-328;342-351;466-473 <S1>
C:Genetics:
A:Gene: GDB:ITGAV; VNRA
A:Cross-references: GDB:120491; OMIM:193210
A:Map position: 2q31-q32
C:Superfamily: integrin alpha-2b chain
C:Keywords: cell adhesion; duplication; glycoprotein; heterodimer; receptor; transmembrane
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1048/Product: integrin alpha-V chain #status experimental <MAT>
F:31-992/Domain: extracellular #status predicted <EXT>
F:993-1016/Domain: transmembrane #status predicted <TMN>
F:1017-1048/Domain: intracellular #status predicted <INT>

Query Match 43.9%; Score 50; DB 2; Length 1048;
Best Local Similarity 38.1%; Pred. No. 70;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPBEKREKLE 21
DB 1017 RMGFPRVRPQEEREQLQ 1037

RESULT 10
TI2766
probable lipoprotein yokB - Bacillus subtilis phage SPBC2

C;Species: *Bacillus subtilis* phage SPBc2
 C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
 C;Accession: T12766; F69907
 R;Lazarevic, V.; Duesterhoeft, A.; Solido, B.; Hilbert, H.; Mael, C.; Karamata, D.
 submitted to the EMBL Data Library, August 1997
 A;Description: The complete nucleotide sequence of the *Bacillus subtilis* SPBc2 prophage
 A;Reference number: Z17583
 A;Accession: T12766
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-236 <LAZ>
 A;Cross-references: UNIPROT:O64016; EMBL:AF020713; NID:G3025478; PID:G3025480; PIDN:AACT
 R;Kunze, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Etrington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
 A;Authors: Schleier, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A;Reference number: A69580; MUID:98044033; PMID:9384377
 A;Accession: F69907
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-236 <KUN>
 A;Cross-references: GB:Z99115; GB:AL009126; NID:G2634478; PIDN:CAB14083.1; PID:e1183612;
 A;Experimental source: strain 168
 C;Genetics:
 A;Gene: yokB

Query Match 43.4%; Score 49.5; DB 2; Length 236;
 Best Local Similarity 84.6%; Pred. No. 19;
 Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
 Qy 8 KKI-PEEKREK 19
 ||| |||||
 Db 205 KKVSPEEKREK 217

RESULT 11
 B60017
 outer capsid protein VP2 - bluetongue virus (serotype 3, strain South Africa-VACC)
 C;Species: bluetongue virus
 C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
 C;Accession: B60017
 R;Gould, A.R.; Pritchard, L.I.
 Virus Res. 17, 31-52, 1990
 A;Title: Relationships amongst bluetongue viruses revealed by comparisons of capsid and
 A;Reference number: A60017; MUID:91021485; PMID:2171239
 A;Accession: B60017
 A;Molecule type: genomic RNA
 A;Residues: 1-959 <GOU>
 A;Cross-references: UNIPROT:Q06998; GB:X55801; NID:G297130; PIDN:CAA39323.1; PID:G297132
 C;Genetics:
 A;Map position: segment 2
 A;Superfamily: bluetongue virus VP2 protein
 C;Keywords: capsid protein; glycoprotein
 F;749,910/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 43.4%; Score 49.5; DB 1; Length 959;
 Best Local Similarity 57.1%; Pred. No. 75;
 Matches 12; Conservative 1; Mismatches 3; Indels 5; Gaps 1;
 Qy 2 LGFFAHKKIPPEEKREKLEQ 22
 ||| |||||
 Db 791 LNFF-----PSYKREKLE 806

RESULT 12

AC1466
 B. subtilis yacL protein homolog lin0266 [imported] - *Listeria innocua* (strain Clip11262)
 C;Species: *Listeria innocua*
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
 C;Accession: AC1466
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
 ok, C.; Schluteter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 C.; Title: Comparative genomics of *Listeria* species.
 A;Reference number: AB1077; MUID:21537279; PMID:11679669
 A;Accession: AC1466
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-357 <GLA>
 A;Cross-references: UNIPROT:Q92F41; GB:AL592022; PIDN:CAC95499.1; PID:gl6412695; GSPDB:G
 A;Experimental source: strain Clip11262
 C;Genetics:
 C;Superfamily: conserved hypothetical protein yacL

Query Match 43.0%; Score 49; DB 2; Length 357;
 Best Local Similarity 71.4%; Pred. No. 33;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 6 AHKKIPPEEKREK 19
 ||| |||||
 Db 146 AKKTPPEEKREK 159

RESULT 13

T06108
 hypothetical protein TSJ17.190 - *Arabidopsis thaliana*
 C;Species: *Arabidopsis thaliana* (mouse-ear cress)
 C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
 C;Accession: T06108
 R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, March 1999
 A;Reference number: Z15184
 A;Accession: T06108
 A;Molecule type: DNA
 A;Residues: 1-615 <BEV>
 A;Cross-references: UNIPROT:Q9SMQ3; EMBL:AL035708; GSPDB:GN00062; ATSP:TSJ17.190
 A;Experimental source: cultivar Columbia; BAC clone TSJ17
 C;Genetics:
 A;Gene: ATSP:TSJ17.190
 A;Map position: 4

Query Match 43.0%; Score 49; DB 2; Length 615;
 Best Local Similarity 71.4%; Pred. No. 56;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 8 KKIPPEEKREKLE 21
 ||| |||||
 Db 396 KKIEKEKKEKLE 409

RESULT 14

S74279
 hypothetical protein YCJ061c - yeast (*Saccharomyces cerevisiae*)
 N;Alternate names: hypothetical protein YCJ060c
 C;Species: *Saccharomyces cerevisiae*
 C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C;Accession: S74279; S19392; S19391; S29373; S21360
 R;Voet, M.; Volckaert, G.
 submitted to the Protein Sequence Database, September 1996
 A;Reference number: S74277
 A;Accession: S74279

Search completed: April 6, 2005, 12:17:14
Job time : 1.76664 secs

```

Query Match      43.0%; Score 49; DB 2; Length 853;
Best Local Similarity 56.2%; Pred. No. 78;
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps

```

RESULT 15
 C71956
 probable endopeptidase Clp ATP-binding chain - Helicobacter pylori (strain J99)
 N;Alternate names: ATP-dependent Clp proteinase regulatory chain
 N;Contains: adenosinetriphosphatase (EC 3.6.1.3)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
 C;Accession: C71956
 R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A;Reference number: A71800; MUID:99120557; PMID:9923692
 A;Accession: C71956
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-856 <ARN>
 A;Cross-references: UNIPROT:Q9ZMH1; GB:AE001462; GB:AE001439; NID:g4154760; PIDN:AAD0582
 A;Experimental source: strain J99
 C;Genetics:
 A;Gene: clpB
 C;Function:
 A;Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon
 e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller
 C;Superfamily: endopeptidase Clp ATP-binding chain
 C;Keywords: ATP; hydrolase; molecular chaperone; nucleotide binding; P-loop
 F;204-211/Region: nucleotide-binding motif A (P-loop)
 F;271-276/Region: nucleotide-binding motif B
 F;605-612/Region: nucleotide-binding motif A (P-loop)
 F;673-678/Region: nucleotide-binding motif B
 F;210/Binding site: ATP (Lys) #status Predicted
 F;611/Binding site: ATP (Lys) #status Predicted

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 3.25607 Seconds
(without alignments)
3459.921 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFPAHKIPBEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	114	100.0	1167	1	ITAG_HUMAN	O75578 homo sapien
2	88	77.2	288	2	QBEM12	QBEM12 mus musculus
3	55	48.2	297	2	Q7PXX0	Q7PXX0 anopheles g
4	53	46.5	214	1	SMP_SALTI	Q820T9 salmonella
5	53	46.5	214	1	SMP_SALTY	Q82JVO salmonella
6	53	46.5	492	2	Q73FL3	Q73FL3 wolbachia p
7	53	46.5	1033	2	Q42598	Q42598 xenopus lae
8	52	45.6	1021	2	Q9GSF4	Q9GSF4 podocoryne
9	51	44.7	128	1	YTE1_BACAN	Q9YRZ2 bacillus an
10	51	44.7	134	2	Q61ED6	Q61ED6 ovis aries
11	51	44.7	180	2	Q8TZJ6	Q8TZJ6 pyrococcus
12	51	44.7	672	2	Q28247	Q28247 archaeoglob
13	51	44.7	856	1	CLPB_HELPY	P71404 helicobacte
14	51	44.7	1013	2	Q8BLA0	Q8BLA0 mus musculus
15	51	44.7	1045	2	Q91292	Q91292 pleurodeles
16	51	44.7	1914	2	Q68FDO	Q68FDO mus musculus
17	51	44.7	1957	2	Q91YC9	Q91YC9 mus musculus
18	50	43.9	145	2	Q6X3V9	Q6X3V9 bacillus th
19	50	43.9	145	2	Q7WSF9	Q7WSF9 bacillus th
20	50	43.9	175	2	Q80Y67	Q80Y67 mus musculus
21	50	43.9	268	2	Q61MW4	Q61MW4 methanococc
22	50	43.9	357	1	Y234_LISMO	Q48762 listeria mo
23	50	43.9	357	2	Q724H8	Q724H8 listeria mo
24	50	43.9	403	2	Q9XYF8	Q9XYF8 trypanosoma
25	50	43.9	505	1	SPKD_SYNY3	P54735 synechocyst
26	50	43.9	1034	1	ITAV_CHICK	P26008 gallus gall
27	50	43.9	1044	1	ITAV_MOUSE	P43406 mus musculus
28	50	43.9	1047	1	ITAV_BOVIN	P80746 bos taurus
29	50	43.9	1048	1	ITAV_HUMAN	P06756 homo sapien
30	49.5	43.4	236	2	Q64016	Q64016 bacterioph
31	49.5	43.4	236	2	Q32005	Q32005 bacillus su

RESULT 1

ID	ITAG_HUMAN	STANDARD;	PRT;	1167 AA.
AC	O75578; Q9UH28;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Integrin alpha-10 precursor.			
GN	Name=ITGAL0;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Articular chondrocytes;			
RX	MEDLINE=98352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;			
RA	Camper L., Hellman U., Lundgren-Akerlund E.,			
RT	"Isolation, cloning, and sequence analysis of the integrin subunit			
RT	alpha10, a beta1-associated collagen binding integrin expressed on			
RT	chondrocytes.";			
RL	J. Biol. Chem. 273:20383-20389(1998).			
RN	(2)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Endothelial cells, and Heart;			
RX	MEDLINE=20169197; PubMed=10702680;			
RA	Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,			
RA	Wang S.-X., Langley R., Krissansen G.W.;			
RT	"The integrin alpha10 subunit: expression pattern, partial gene			
RT	structure, and chromosomal localization.";			
RL	Cytogenet. Cell Genet. 87:238-244(1999).			
CC	-!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.			
CC	-!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10			
CC	associates with beta-1.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: Widely expressed with highest expression in			
CC	muscle and heart. Found in articular cartilage.			
CC	-!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins			
CC	with I-domains do not undergo protease cleavage.			
CC	-!- SIMILARITY: Belongs to the integrin alpha chain family.			
CC	-!- SIMILARITY: Contains 7 FG-GAP repeats.			
CC	-!- SIMILARITY: Contains 1 VWFA domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF074015; AAC31952.1; -			
DR	EMBL; AF112345; AAF21944.1; -			
DR	EMBL; AF172723; AAF61638.1; -			

ALIGNMENTS

32	49.5	43.4	635	2	Q96LR8	Q96LR8 homo sapien
33	49.5	43.4	846	2	Q86W98	Q86W98 homo sapien
34	49.5	43.4	959	1	VP2_BT3V3V	Q69598 bluetongue
35	49.5	43.4	959	2	Q6RYD5	Q69598 bluetongue
36	49.5	43.4	1372	2	Q9HAW1	Q9HAW1 homo sapien
37	49.5	43.4	1388	2	Q9HAW2	Q9HAW2 homo sapien
38	49.5	43.4	1985	2	Q7RC59	Q7RC59 plasmodium
39	49.5	43.4	2187	2	Q9H197	Q9H197 homo sapien
40	49.5	43.4	2254	2	Q9HCY0	Q9HCY0 homo sapien
41	49.5	43.4	2627	2	Q9C0H4	Q9C0H4 homo sapien
42	49	43.0	357	1	Y266_LISIN	Q92F41 listeria in
43	49	43.0	615	2	Q9SMQ3	Q9SMQ3 arabidopsis
44	49	43.0	856	1	CLPB_HELPJ	Q9ZMH1 helicobacte
45	49	43.0	1096	1	MRC1_YEAST	P25588 saccharomyc

DR HSP; P18614; IMHP.
 DR GENE; HGNC:6135; ITGA10.
 DR MIM; 604042; .
 DR GO; GO:0008305; C:integrin complex; TAS.
 DR GO; GO:000518; P:collagen binding; TAS.
 DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 3.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PRO1185; INTEGRIN.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
 DR PROSITE; PS00234; VWFA; 1.
 DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
 KW Repeat; Signal; Transmembrane.
 FT SIGNAL 1 22 Potential.
 FT CHAIN 23 1167 Integrin alpha-10.
 FT DOMAIN 23 1122 Extracellular (Potential).
 FT TRANSMEM 1123 1145 Potential.
 FT DOMAIN 1146 1167 Cytoplasmic (Potential).
 FT REPEAT 38 97 FG-GAP 1.
 FT REPEAT 98 ? FG-GAP 2.
 FT DOMAIN 167 350 VWFA.
 FT REPEAT 365 427 FG-GAP 3.
 FT REPEAT 428 482 FG-GAP 4.
 FT REPEAT 483 545 FG-GAP 5.
 FT REPEAT 546 605 FG-GAP 6.
 FT REPEAT 608 660 FG-GAP 7.
 FT DOMAIN 1134 1140 Poly-Leu.
 FT CA_BIND 494 502 Potential.
 FT CA_BIND 558 566 Potential.
 FT CA_BIND 620 628 Potential.
 FT DISULFID 76 86 By similarity.
 FT DISULFID 666 675 By similarity.
 FT DISULFID 681 736 By similarity.
 FT DISULFID 789 795 By similarity.
 FT CARBOHYD 98 98 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 234 234 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 336 336 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 364 364 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 733 733 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 763 763 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 839 839 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 921 921 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1011 1011 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1018 1018 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1039 1039 N-linked (GlcNAc...) (Potential).
 FT CONFLICT 844 844 I -> L (in Ref. 2).
 FT CONFLICT 909 909 G -> V (in Ref. 2).
 FT CONFLICT 926 926 E -> D (in Ref. 2).
 SQ SEQUENCE 1167 AA; 127573 MW; A87D3A1C25C1AE0 CRC64;

Query Match 100.0%; Score 114; DB 1; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 8.3e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPEEKREKLEQ 22
 |||||
 Db 1146 KLGFPAHKKIPEEKREKLEQ 1167

RESULT 2
 Q8BM12 PRELIMINARY; PRT; 288 AA.
 AC Q8BM12

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus 6 days neonate skin cDNA, RIKEN full-length enriched
 library, clone:A030011P23 product:INTEGRIN ALPHA-10 homolog

DE (Fragment).
 GN Name=Itgal0;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX The FANTOM Consortium;
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
 Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hatada A.,
 Yamanoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 Fukuda S., Furuno M., Hanagaki T., Hata A., Hashizume W.,
 Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 Hori F., Imotani K., Ishii Y., Itoh M., Kondo H., Kouda M., Koya S.,
 Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).
 -!- SIMILARITY: Belongs to the integrin alpha chain family.
 DR EMBL; AK037222; BAC29761.1; .
 DR MGI; MGI:2153482; Itgal0.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0008305; C:integrin complex; IEA.
 DR GO; GO:0005515; F:protein binding; IEA.

```

DR GO: GO:0007160; P:cell-matrix adhesion; IEA.
DR GO: GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF00357; Integrin_alpha; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN_1.
KW Cell adhesion; Integrin; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 288 AA; 31949 MW; 3F6200F9C9475E85 CRC64;

Query Match
Best Local Similarity 77.2%; Score 88; DB 2; Length 288;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPPEEKREKLEQ 22
Db 267 KLGFFTTRKKIPKEESEKLEQ 288

RESULT 3
ID Q7PXX0 PRELIMINARY; PRT; 997 AA.
AC Q7PXX0;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AGCP12053 (Fragment)
GN Name=agCS6510; ORFNames=ENSANGG00000014880;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008987; EAA00964.1; -.
DR InterPro; IPR007110; IG-like.
DR PROSITE; PS50835; IG_LIKE; 6.
FT NON_TER 1
FT NON_TER 997
FT NON_TER 997
SQ SEQUENCE 997 AA; 113756 MW; 3C3110C77B646F94 CRC64;

Query Match
Best Local Similarity 48.2%; Score 55; DB 2; Length 997;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 8 KKIPPEEKREKLE 21
Db 840 KKVPEEKREKVE 853

RESULT 4
ID SMP_SALTI STANDARD; PRT; 214 AA.
AC Q8ZT09;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Protein smp precursor.
GN Name=smp; OrderedLocustNames=STY4924, t4616;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RA MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,

```

```

RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jegels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RX DOI=10.1128/JB.185.7.2330-2337.2003;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1- SIMILARITY: Belongs to the smp family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL627284; CAD03408.1; -.
DR EMBL; AE016849; AA072048.1; -.
KW Complete proteome; Membrane; Signal.
FT SIGNAL 1 30 Potential.
FT CHAIN 31 214 Protein smp.
SQ SEQUENCE 214 AA; 24279 MW; 2683D0DF9174BF80 CRC64;

Query Match
Best Local Similarity 46.5%; Score 53; DB 1; Length 214;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 6 AHKKIPPEEKREK 19
Db 200 ASKPVPPEEREK 213

RESULT 5
ID SMP_SALTY STANDARD; PRT; 214 AA.
AC Q8ZTVO;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Protein smp precursor.
GN Name=smp; OrderedLocustNames=STM4577;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L72 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RX Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).

```

DT	01-JAN-1998	(TrEMBLrel. 05, Last sequence update)
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)
DE	Integrin alphav subunit.	
OS	Xenopus laevis (African clawed frog).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;	
OC	Xenopodinae; Xenopus.	
OX	NCBI_TaxID=8355;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=98211587; PubMed=9551862;	
RT	Joo S.T.O., Reintsch W.E., Brinker A., Klein C., Hausen P.;	
RA	"Cloning of the Xenopus integrin alpha(v) subunit and analysis of its	
RL	distribution during early development.";	
RL	Int. J. Dev. Biol. 42:171-179(1998).	
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).	
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.	
DR	EMBL; U92006; AABG2090.1; -.	
DR	HSP; P06756; IL5G.	
DR	GO; GO:0016021; C:integral to membrane; IEA.	
DR	GO; GO:0008305; C:integrin complex; IEA.	
DR	GO; GO:0005515; P:protein binding; IEA.	
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.	
DR	GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.	
DR	InterPro; IPR000413; Integrin_alpha.	
DR	Pfam; PF00357; Integrin_alpha; 1.	
DR	PRINTS; PR01185; INTEGRIN_A.	
DR	SMART; SM00191; Int_alpha; 5.	
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.	
KW	Cell adhesion; Integrin; Transmembrane.	
SQ	SEQUENCE 1033 AA; 114024 MW; 35F73B0765C06BD6 CRC64;	
	Query Match 46.5%; Score 53; DB 2; Length 1033;	
	Best Local Similarity 47.8%; Pred. NO. 1.6e+02;	
	Matches 10; Conservative 7; Mismatches 7; Indels 0; Gaps	
Qy	1 KLGFFAHKKIPKEEKREKKLE 21	
	: :	
Dd	1002 KLGFPPKVRPPQEETEREQLQ 1022	
	: :	
RESULT 8		
ID	OQSFP4 PRELIMINARY; PRT; 1021 AA.	
AC	OQSFP4;	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	Integrin alpha chain.	
GN	Name=INTA;	
OS	Podocoryne carnea.	
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;	
OC	Hydractiniidae; Podocoryne.	
OX	NCBI_TaxID=6096;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21375962; PubMed=11482899; DOI=10.1006/cbir.2000.0708;	
RT	Reber-Muller S., Studer R., Muller P., Yanze N., Schmid V.;	
RA	"Integrin and talin in the Jellyfish Podocoryne carnea.";	
RL	Cell Biol. Int. 25:753-769(2001).	
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.	
DR	EMBL; AF308251; AAG25993.1; -.	
DR	GO; GO:0016021; C:integral to membrane; IEA.	
DR	GO; GO:0008305; C:integrin complex; IEA.	
DR	GO; GO:0005515; P:protein binding; IEA.	
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.	
DR	GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.	
DR	InterPro; IPR000413; Integrin_alpha.	
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.	
DR	PRINTS; PR01185; INTEGRIN_A.	
DR	SMART; SM00191; Int_alpha; 5.	
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.	
KW	Cell adhesion; Integrin; Transmembrane.	
SQ	SEQUENCE 1033 AA; 114024 MW; 35F73B0765C06BD6 CRC64;	
	Query Match 46.5%; Score 53; DB 2; Length 1033;	
	Best Local Similarity 47.8%; Pred. NO. 1.6e+02;	
	Matches 10; Conservative 7; Mismatches 7; Indels 0; Gaps	

```

DR PRINTS; PRO1185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN_1.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 1021 AA; 110986 MW; 28A8E782FF616DB2 CRC64;

Query Match 45.6%; Score 52; DB 2; Length 1021;
Best Local Similarity 55.6%; Pred. No. 2.2e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KLGFFAHKKIPEEKREE 18
Db 996 KKGFFKKKGWGEDEEE 1013

RESULT 9
YTEL BACAN
ID YTEL BACAN STANDARD; PRT; 128 AA.
AC Q9RMZ2
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Hypothetical protein pX02-40/EXB0041/GBAA_pX02_0041.
GN OrderedLocusNames=pX02-40, EXB0041, GBAA_pX02_0041;
OS Bacillus anthracis.
OG Plasmid pX02.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pasteur;
RX PubMed=10475962;
RA Okihara R.T., Cloud K., Hampton O., Hoffmaster A., Hill K.K., Keim P., Koehler T., Lamke G., Kumano S., Manter D., Martinez Y., Ricke D., Svensson R., Jackson P.J.;
RA "Sequence, assembly and analysis of pX01 and pX02.";
RL J. Appl. Microbiol. 87:261-262(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate Florida / A2012;
RX MEDLINE=22061436; PubMed=12004073; DOI=10.1126/science.1071837;
RA Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L., Holtzapfel E., Busch J.D., Smith K.L., Schupp J.M., Solomon D., Keim P., Fraser C.M.;
RA "Comparative genome sequencing for discovery of novel polymorphisms in Bacillus anthracis.";
RL Science 296:2028-2033(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581;
RA Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B., Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L., Fraser C.M.;
RA "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ebi.ac.uk/announcements/ or send an email to license@ebi.ac.uk).

```

```

RESULT 12
O28247 ID O28247 PRELIMINARY; PRT; 672 AA.
AC O28247;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Signal-transducing histidine kinase, putative.
GN OrderedLocusNames=AF2032;
OS Archaeoglobus fulgidus.
OS Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475; DOI=10.1038/37052;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Karlvage A.R., Graham D.H., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,
RA Peterson S.N., Reich C.I., McNeil L.K., Badger J.H., Glodek A.,
RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
RA Uterback T.R., Cotton M.D., Spriggs T., Artach P., Kaine B.P.,
RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,
RA Woese C.R., Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE000963; AA889224.1; -.
DR PIR; G69503; G69503.
DR TIGR; AF2032; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. .; IEA.
DR InterPro; IPR003661; His_kin_N.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS.
DR InterPro; IPR000700; PAS-assoc_C.
DR Pfam; PF00512; HisKA; 1.
DR Pfam; PF00785; PAC; 3.
DR Pfam; PF00989; PAS; 3.
DR SMART; SM00388; HisKA; 1.
DR SMART; SM00086; PAC; 3.
DR SMART; SM00091; PAS; 4.
DR TIGRFAMs; TIGR00229; sensory_box; 4.
DR PROSITE; PS50113; PAC; 3.
DR PROSITE; PS50112; PAS; 3.
KW Complete proteome; Kinase.
SQ SEQUENCE 672 AA; 78620 MW; F95DC675BBD8F749 CRC64;

Query Match 44.7%; Score 51; DB 2; Length 672;
Best Local Similarity 55.6%; Pred No. 2e+02;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 5 FAHKKIPEEKREKRLQ 22
Db 572 FAPKSMEDERRELLAQ 589
||| : :||| :|
||| : :||| :|

RESULT 13
CLPB_HELPY
ID CLPB_HELPY STANDARD; PRT; 856 AA.
AC P71404;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Chaperone clpB.

```

Name=clpB; OrderedLocusNames=HP0264;
 Helicobacter pylori (Campylobacter pylori).
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11637;
 RX PubMed=9440536;
 RA Allan E., Mullany P., Tabaqchali S.;
 RT "Construction and characterization of a Helicobacter pylori clpB
 RT mutant and role of the gene in the stress response.";
 RL J. Bacteriol. 180:426-429(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185; DOI=10.1038/41483;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S.R.,
 RA Dougherty B.A., Nelson K.E., Quackenbush J., Zhou L., Kirkness E.F.,
 RA Peterson S.N., Loftus B.J., Richardson D.L., Dodson R.J., Khalak H.G.,
 RA Glodek A., McKenney K., FitzGerald L.M., Lee N., Adams M.D.,
 RA Hickey E.K., Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D.,
 RA Kelley J.M., Cotton M.D., Weidman J.F., Fujii C., Bowman C.,
 RA Wathey L., Wallin E., Hayes W.S., Borodovsky M., Karp P.D.,
 RA Smith H.O., Fraser C.M., Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori.";
 RL Nature 388:539-547(1997).
 CC -1- FUNCTION: Part of a stress-induced multi-chaperone system, it is
 CC involved in the recovery of the cell from heat-induced damage, in
 CC cooperation with dnaK, dnaJ and grpE. Acts before dnaK, in the
 CC processing of protein aggregates. Protein binding stimulates the
 CC ATPase activity; ATP hydrolysis unfolds the denatured protein
 CC aggregates, which probably helps expose new hydrophobic binding
 CC sites on the surface of clpB-bound aggregates, contributing to the
 CC solubilization and refolding of denatured protein aggregates by
 CC dnaK (By similarity). Necessary for surviving high-temperature
 CC stress.
 CC -1- SUBUNIT: Homohexamer. The oligomerization is ATP-dependent (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- DOMAIN: The N-terminal domain probably functions as a substrate-
 CC discriminating domain, recruiting aggregated proteins to the clpB
 CC hexamer and/or stabilizing bound proteins. The NBD2 domain is
 CC responsible for oligomerization, whereas the NBD1 domain
 CC stabilizes the hexamer probably in an ATP-dependent manner. The
 CC movement of the coiled-coil domain is essential for clpB ability
 CC to rescue proteins from an aggregated state, probably by pulling
 CC apart large aggregated proteins, which are bound between the
 CC coiled-coils motifs of adjacent clpB subunits in the functional
 CC hexamer (By similarity).
 CC -1- SIMILARITY: Belongs to the clpA/clpB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y08238; CAA69406.1; -.
 CC EMBL; AE000545; AAD07330.1; -.
 CC PIR; H64552; H64552.
 CC HSP; P03815; IJBK.
 CC TIGR; HP0264; -.
 CC InterPro; IPR003593; AAA_ATPase.
 CC InterPro; IPR003959; AAA_ATPase_cent.
 CC InterPro; IPR001270; Chaperin_clpA/B.
 CC InterPro; IPR004176; Clp_N.
 CC Pfam; PF00004; AAA; 2.
 CC Pfam; PF02861; Clp_N; 2.

```
DR PRINTS; PRO0300; CLPPROTEASEA.
DR SMART; SMO0382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding; Chaperone; Coiled coil; Complete proteome; Heat shock;
KW Repeat.
FT FT DOMAIN 1 141 N-terminal (By similarity).
FT FT DOMAIN 157 337 NBD1 (By similarity).
FT FT DOMAIN 338 545 Linker (By similarity).
FT FT DOMAIN 555 763 NBD2 (By similarity).
FT FT DOMAIN 764 856 C-terminal (By similarity).
FT FT DOMAIN 388 522 Coiled coil (By similarity).
FT FT NP_BIND 204 211 ATP 1 (By similarity).
FT FT NP_BIND 605 612 ATP 2 (By similarity).
FT FT CONFLICT 14 14 T -> A (in Ref. 1).
FT FT CONFLICT 33 33 M -> L (in Ref. 1).
FT FT CONFLICT 59 59 Q -> E (in Ref. 1).
FT FT CONFLICT 62 62 R -> K (in Ref. 1).
FT FT CONFLICT 79 79 S -> N (in Ref. 1).
FT FT CONFLICT 103 103 R -> T (in Ref. 1).
FT FT CONFLICT 119 119 G -> S (in Ref. 1).
FT FT CONFLICT 131 131 A -> T (in Ref. 1).
FT FT CONFLICT 145 145 R -> A (in Ref. 1).
FT FT CONFLICT 149 149 D -> G (in Ref. 1).
FT FT CONFLICT 222 222 M -> V (in Ref. 1).
FT FT CONFLICT 435 435 A -> H (in Ref. 1).
FT FT CONFLICT 465 465 V -> A (in Ref. 1).
FT FT CONFLICT 643 644 MS -> IT (in Ref. 1).
FT FT CONFLICT 649 649 A -> P (in Ref. 1).
FT FT CONFLICT 734 734 E -> D (in Ref. 1).
FT FT CONFLICT 765 765 D -> G (in Ref. 1).
FT FT CONFLICT 836 836 V -> I (in Ref. 1).
FT FT CONFLICT 840 840 D -> G (in Ref. 1).
SQ SEQUENCE 856 AA; 96683 MW; E3902C7D989496AD CRC64;

Query Match 44.7%; Score 51; DB 1; Length 856;
Best Local Similarity 45.0%; Pred. No. 2.5e+02;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHKPIPEEKKEEKLEQ 22
Db 496 GEIYSKIPENKKEELQR 515

RESULT 14
Q8BLA0
ID Q8BLA0 PRELIMINARY; PRT; 1013 AA.
AC Q8BLA0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
DE enriched library, clone:B230314P12 product:similar to TUMOR SUPPRESSOR
DE P53-BINDING PROTEIN 1 (P53-BINDING PROTEIN 1) (53BP1) (Fragment).
GN Name=trp53bp1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
```

```
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
Hori F., Imotani K., Ishii Y., Itoh M., Kigawa I., Kouda M., Koya S.,
Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Nakamura M.,
Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
Nishi K., Nomura K., Numazaki R., Ohno N., Ohsato N., Okazaki Y.,
Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
Tegawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK045851; BAC32512.1; -.
DR MGD; MGI:1351320; Trp53bp1.
DR GO; GO:0000776; C:kinetochore; IDA.
DR GO; GO:0003684; F:damaged DNA binding; IDA.
DR GO; GO:0008134; F:transcription factor binding; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
FT NON TER 1013 1013
SQ SEQUENCE 1013 AA; 109079 MW; 4AAE3FDA60E42F72 CRC64;

Query Match 44.7%; Score 51; DB 2; Length 1013;
Best Local Similarity 69.2%; Pred. No. 3e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 10 IPPEEKKEEKLEQ 22
Db 129 LPPEEKKEEKLEQ 141

RESULT 15
Q91292
ID Q91292 PRELIMINARY; PRT; 1045 AA.
AC Q91292;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Integrin.
OS Pleurodeles waltlii (Iberian ribbed newt).
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
 OC Pleurodeles.
 OX NCBI_TaxID=8319;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95377519; PubMed=7649360; DOI=10.1006/dbio.1995.1212;
 RA Alfandari D., Whittaker C.A., Desimone D.W., Darribere T.;
 RT "Integrin alpha v subunit is expressed on mesodermal cell surfaces
 during amphibian gastrulation.";
 RL Dev. Biol. 170:249-261(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Alfandari D.R.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
 CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
 DR EMBL; X81108; CAAS7014.1; -.
 DR PIR; S60571; S60571.
 DR HSP; P06756; IL5G.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0008305; C:integrin complex; IEA.
 DR GO; GO:000515; P:protein binding; IEA.
 DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
 DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
 DR Pfam; PF00357; Integrin alpha; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR SMART; SM00191; Int_alpha; 5.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN_1.
 KW Cell adhesion; Integrin; Transmembrane.
 SQ SEQUENCE 1045 AA; 115421 MW; 4DB34B766B0C648E CRC64;

Query Match 44.7%; Score 51; DB 2; Length 1045;
 Best Local Similarity 42.9%; Pred.No. 3.1e+02;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 1 KLGFFAHKKIPESEKREEKLE 21
 ||||| : ||| : |||
 Db 1014 KFGFFKRVPPQEQREQLQ 1034

Search completed: April 6, 2005, 12:15:39
 Job time : 5.25607 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 2.7823 Seconds
(without alignments)
3058.161 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114

Sequence: 1 KLGFPAKKIPBEKREKLEQ 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	100.0	22	3	AAY32244 Human int
2	114	100.0	1049	5	ADR41424 Human CD-
3	114	100.0	1132	3	AAY32243 Human int
4	114	100.0	1167	3	AAY32242 Human int
5	114	100.0	1167	4	AAB64584 Human sec
6	114	100.0	1167	6	ABP99490 Human sec
7	114	100.0	1167	6	ABR00964 Human gen
8	114	100.0	1167	6	ADA44026 Human sec
9	114	100.0	1167	8	ADQ19290 Human sof
10	52	45.6	344	7	ADC97404 E. faeciu
11	51	44.7	121	4	ABG26508 Novel hum
12	51	44.7	147	4	ABG11269 Novel hum
13	51	44.7	502	4	ABG07742 Novel hum
14	51	44.7	856	4	AAU35717 Helicobac
15	51	44.7	856	6	ABU30742 Protein e
16	50	43.9	32	2	AAR80955 Integrin
17	50	43.9	156	4	ABG18894 Novel hum
18	50	43.9	357	5	ABR48958 Listeria
19	50	43.9	1048	5	AAU76335 Human ant
20	50	43.9	1048	6	ABU03549 Angiogene
21	50	43.9	1048	6	AAO27098 Human int
22	50	43.9	1048	6	ABB82767 Human alp
23	50	43.9	1048	7	ADD48891 Human Pro
24	50	43.9	1061	4	ABG18895 Novel hum
25	49.5	43.4	1388	6	ABJ38696 Human nuc

26	49.5	43.4	1398	7	ADJ69333 Human hea
27	49	43.0	130	4	AAO0651 Human pol
28	49	43.0	413	5	ADH32817 Yeast sMO
29	49	43.0	486	3	AAG53097 Arabidops
30	49	43.0	584	3	AAG53096 Arabidops
31	49	43.0	615	3	AAG53095 Arabidops
32	49	43.0	751	2	AAW13491 Helicobac
33	49	43.0	856	4	AAU35896 Helicobac
34	48.5	42.5	407	8	ADP49328 S Pyrogen
35	48.5	42.5	723	4	ABG28392 Novel hum
36	48	42.1	573	2	AAR82026 Squalene
37	48	42.1	573	7	ADB85174 Rat equal
38	48	42.1	1464	4	ABB61199 Drosophil
39	48	42.1	1526	7	ADB61230 Rat Prote
40	48	42.1	1526	7	ADD44993 Rat Prote
41	47.5	41.7	2343	2	AAW80989 Canine fa
42	47.5	41.7	2343	3	AAY57846 Canine fa
43	47	41.2	136	4	ABG14329 Novel hum
44	47	41.2	177	3	AAG53177 Arabidops
45	47	41.2	177	3	AAG14466 Arabidops

ALIGNMENTS

RESULT 1

AAAY32244

ID AAY32244 standard; peptide; 22 AA.

XX AC AAY32244;

XX AC AAY32244;

DT 15-FEB-2000 (first entry)

XX Human integrin subunit alpha-10 cytoplasmic domain peptide.

DE Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;

XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;

KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;

KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.

XX Homo sapiens.

OS Homo sapiens.

XX WO9951639-A1.

PN 14-OCT-1999.

PD 31-MAR-1999; 99WO-SE000544.

XX 02-APR-1998; 98SE-00001164.

XX 28-JAN-1999; 99SE-00000319.

XX (ACTI-) ACTIVE BIOTECH AB.

XX Lundgren-Akerlund E;

XX WPI; 2000-052639/04.

XX New isolated integrin subunit alpha-10, used as a marker or target

PT molecule for cells during development, regeneration and pathological

PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or

PT inflammation.

XX Claim 21; Page 53; 90pp; English.

XX This sequence represents a fragment of novel human chondrocyte integrin

XX subunit alpha-10 (ISa10, see AAY32242), corresponding to the C-terminal

XX cytoplasmic domain of the protein. The invention relates to a recombinant

CC or isolated integrin heterodimer comprising the alpha10 subunit in

CC association with subunit beta (especially beta-1). The integrin

CC heterodimer, or the subunit alpha-10, or a fragment of it such as the

CC present sequence, can be used as a marker or target of all types of

CC cells, e.g. of chondrocytes, osteoblasts and fibroblasts. They can also

CC be used: for treating pathological conditions involving ISa10, such as

CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis; for

CC detecting the formation of cartilage during embryonal development,
 CC physiological or therapeutic regeneration of cartilage, or detecting
 CC regeneration of cartilage or chondrocytes during transplantation of
 CC cartilage or chondrocytes; for selection and analysis or for sorting,
 CC isolating or purification of chondrocytes and for in vitro studies of
 CC differentiation of chondrocytes; and as a target for anti-adhesive drugs
 CC or molecules in tendon, ligament, skeletal muscle or other tissues where
 CC adhesion impairs the function of the tissue (all claimed). ISa10 binding
 CC entities can be used to determine the differentiation-state of cells
 CC during embryonic development, angiogenesis or development of cancer, in
 CC pathological conditions such as rheumatoid arthritis, osteoarthritis or
 CC cancer, in tissue regeneration or in therapeutic and physiological
 CC repair of cartilage (claimed). A vaccine comprising the integrin
 CC heterodimer or subunit alpha-10 is also claimed. ISa10 polynucleotides,
 CC vectors, host cells and methods of producing recombinant ISa10 are also
 CC claimed

XX Sequence 22 AA;
 SQ

Query Match 100.0%; Score 114; DB 3; Length 22;
 Best Local Similarity 100.0%; Pred. No. 5.7e-09;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPEEKREKLEQ 22
 |||||
 Db 1 KLGFPAHKKIPEEKREKLEQ 22

RESULT 2
 ADR41424
 ID ADR41424 standard; protein; 1049 AA.
 XX
 AC ADR41424;
 XX
 DT 07-OCT-2004 (first entry)
 XX
 DE Human CD-like molecule HA0AD2, SEQ ID NO:223.
 XX
 KW Human; CD-like molecule; cluster of differentiation; diagnosis;
 KW prevention; immune disorder; immunodeficiency; autoimmune disorder;
 KW blood-related disorder; haematological disorder; haemostatic disorder;
 KW thrombolytic disorder; hyperproliferative disorder; cancer; tumour;
 KW apoptotic disorder; cardiovascular disorder; respiratory disorder;
 KW oncogenic disorder; neovascularisation; neurological disorder;
 KW endocrine disorder; reproductive system disorder; infectious disease;
 KW gastrointestinal disorder; drug screening; tissue regeneration;
 KW chemotaxis; gene therapy; antibody therapy; drug targeting;
 KW chromosome mapping; forensic analysis; immunophenotyping; cytostatic;
 KW haemostatic; tranquiliser; vulnerary; antiinflammatory; nephrotropic;
 KW cardiant; anti-allergic; anti-HIV; antirheumatic; antiarthritic;
 KW antipneumatic; immunosuppressive; vasotropic; nootropic; neuroprotective;
 KW antithyroid; thyromimetic; gynaecological; virucide; hepatotropic;
 KW antibacterial; dermatological; chromosome iq21.
 XX
 OS Homo sapiens.
 XX
 PN WO200226930-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029838.
 XX
 PR 26-SEP-2000; 2000US-0235484P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 FI Rosen CA, Birse CE;
 XX
 DR WPI; 2002-405050/43.
 DR N-PSDB; ADR41248.
 XX
 PT Novel polynucleotides and polypeptides useful for treating, preventing or
 PT ameliorating cardiovascular, renal, neurovascular, and autoimmune

PT disorders.
 XX
 PS Claim 11; SEQ ID NO 223; 1243pp; English.
 XX
 CC The invention relates to 167 novel human CD (cluster of differentiation)-
 CC like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid11)-
 XX Sequence 1049 AA;
 SQ

Query Match 100.0%; Score 114; DB 5; Length 1049;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPEEKREKLEQ 22
 |||||
 Db 1028 KLGFPAHKKIPEEKREKLEQ 1049

RESULT 3
 AAY32243
 ID AAY32243 standard; protein; 1132 AA.
 XX
 AC AAY32243;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Human integrin subunit alpha-10 splice variant.
 XX
 KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;
 KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker;
 KW splice variant.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..22 /note= "signal peptide"
 FT Protein 23..1132 /note= "mature protein"
 FT
 XX WO9951639-A1.
 XX
 PD 14-OCT-1999.
 XX
 PF 31-MAR-1999; 99WO-SE000544.
 XX
 PR 02-APR-1998; 98SE-00001164.
 PR 28-JAN-1999; 99SE-00000319.
 XX
 PA (ACTI-) ACTIVE BIOTECH AB.
 XX
 FI Lundgren-Akerlund B;
 XX
 DR WPI; 2000-052639/04.
 DR N-PSDB; AAZ34720.
 XX
 PT New isolated integrin subunit alpha-10, used as a marker or target
 PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation.
 XX
 PS Claim 1; Page 43-48; 90pp; English.
 XX
 CC This sequence represents a splice variant of novel human chondrocyte
 CC integrin subunit alpha-10 (ISa10). It is identical to ISa10 (see
 CC AAY32242) except for deletion of amino acids 975-986. The invention
 CC relates to a recombinant or isolated integrin heterodimer comprising the
 CC alpha10 subunit in association with subunit beta (especially beta-1). The
 CC heterodimer, subunit alpha-10 or splice variant can be used as a marker
 CC or target of all types of cells, e.g. of chondrocytes, osteoblasts and
 CC fibroblasts. They can also be used: for treating pathological conditions
 CC involving ISa10, such as damage to cartilage, trauma, rheumatoid

CC arthritis or osteoarthritis; for detecting the formation of cartilage
 CC during embryonal development, physiological or therapeutic repair of
 CC cartilage, or detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and analysis
 CC or for sorting, isolating or purification of chondrocytes and for in
 CC vitro studies of differentiation of chondrocytes; and as a target for
 CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or
 CC other tissues where adhesion impairs the function of the tissue (all
 CC claimed). ISa10 binding entities can be used to determine the
 CC differentiation-state of cells during embryonic development, angiogenesis
 CC or development of cancer, in pathological conditions such as rheumatoid
 CC arthritis, osteoarthritis or cancer, in tissue regeneration or in
 CC therapeutic and physiological repair of cartilage (claimed). A
 CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also
 CC claimed. ISa10 polynucleotides, vectors, host cells and methods of
 CC producing recombinant ISa10 are also claimed
 XX
 SQ Sequence 1132 AA;

Query Match 100.0%; Score 114; DB 3; Length 1132;
 Best Local Similarity 100.0%; Pred. NO. 3e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPEEKREKLEQ 22
 |||||
 Db 1111 KLGFFAHKKIPEEKREKLEQ 1132

RESULT 4
 ID AAY32242
 XX AAY32242 standard; protein; 1167 AA.

AC AAY32242;

XX 15-FEB-2000 (first entry)

DT Human integrin subunit alpha-10.

DE Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;

KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;

KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.

XX Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..22

FT Protein 23..1145

FT Protein 23..1145

FT Domain 23..1120

FT Modified-site 98

FT Domain 162..359

FT Domain 162..359

FT Modified-site 336

FT Modified-site 364

FT Binding-site 494..502

FT Binding-site 558..566

FT Binding-site 620..628

FT Binding-site 733

FT Modified-site 839

FT Modified-site 921

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1018

FT Modified-site 1039

FT Domain 1121..1145

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

FT Domain 1122..1167

DT 22-MAR-2001 (first entry)
 DE Human secreted protein #37.
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 OS Homo sapiens.
 XX WO200077197-A1.
 PN 21-DEC-2000.
 XX 01-JUN-2000; 2000WO-US014934.
 PF 11-JUN-1999; 99US-0138599P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX Rosen CA, Ruben SM, Komatsoulis GA;
 PI WPI; 2001-032312/04.
 DR N-PSDB; AAF32793.
 XX Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 XX Claim 11; Page 496-500; 558pp; English.
 CC Sequences AAB64549-B64594 represent the amino acid sequences of 47 human
 CC secreted proteins encoded by the genes AAF32757-F32803. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from
 CC a range of human tissues disclosed in the specification. The nucleic
 CC acids, proteins, antibodies and (ant)agonists are useful in the
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)
 CC immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections
 XX SQ Sequence 1167 AA;
 Query Match 100.0%; Score 114; DB 4; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLGFPAHKKIPEEKREKLEQ 22
 DB 1146 KLGFPAHKKIPEEKREKLEQ 1167
 RESULT 6
 ID AABP99490 standard; protein; 1167 AA.
 XX AABP99490;
 AC AABP99490;
 XX 26-MAR-2003 (first entry)
 DT Human secreted protein SEQ ID NO 434.
 DE Human; secreted protein; nootropic; neuroprotective; cytostatic;
 KW Human; secreted protein; nootropic; neuroprotective; cytostatic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW vulnary; antibacterial; antiparkinsonian; antisklking; antianemic;
 KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
 KW antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;
 KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 KW gene therapy.
 OS Homo sapiens.
 XX WO200277186-A2.
 PN 03-OCT-2002.
 XX 26-MAR-2002; 2002WO-US009188.
 PF 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM;
 XX WPI; 2003-040583/03.
 DR N-PSDB; ABZ66911.
 XX New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
 PT West Nile fever.
 XX Claim 1; Page 1395-1398; 2423pp; English.
 CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections
 XX SQ Sequence 1167 AA;
 Query Match 100.0%; Score 114; DB 6; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLGFPAHKKIPEEKREKLEQ 22
 DB 1146 KLGFPAHKKIPEEKREKLEQ 1167
 RESULT 7
 ID ABR00964 standard; protein; 1167 AA.
 XX ABR00964;
 AC ABR00964;
 XX 12-MAY-2003 (first entry)
 DT Human gene 18-encoded secreted protein HA0AG15, SEQ ID NO:445.
 DE Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiotenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;

KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnery; chromosome iq21.
 XX
 OS Homo sapiens.
 XX
 PN WO200277013-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-US009370.
 XX
 PR 27-MAR-2001; 2001US-0278650P.
 XX
 PR 12-SEP-2001; 2001US-00950082.
 XX
 PR 12-SEP-2001; 2001US-00950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 PI WPI; 2003-040578/03.
 XX
 DR N-PSDB; ABZ73298.
 XX
 DR New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX
 PS Claim 13; Page 1387-1390; 2474pp; English.
 XX
 CC ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein of the invention
 XX
 SQ Sequence 1167 AA;
 Query Match 100.0%; Score 114; DB 6; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 KLGFFAHKKIPPEEKREKLEQ 22
 |||||
 Db 1146 KLGFFAHKKIPPEEKREKLEQ 1167
 RESULT 8
 ADA44026
 ID ADA44026 standard; protein; 1167 AA.
 XX
 AC ADA44026;
 XX
 XX 20-NOV-2003 (first entry)
 DT
 XX Human secreted protein SEQ ID 214.
 DE
 XX Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;
 KW

KW Neuroprotective; Cerebroprotective; Antianemic.
 XX
 OS Homo sapiens.
 XX
 PN WO2003000865-A2.
 XX
 PD 03-JAN-2003.
 XX
 PF 26-MAR-2002; 2002WO-US009105.
 XX
 PR 27-MAR-2001; 2001US-0278650P.
 XX
 PR 12-SEP-2001; 2001US-00950082.
 XX
 PR 12-SEP-2001; 2001US-00950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 PI WPI; 2003-184045/18.
 XX
 DR N-PSDB; ADA43832.
 XX
 DR A human secreted protein and nucleic acids useful for preparing a
 PT diagnostic or pharmaceutical composition for diagnosing or treating
 PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,
 PT retinopathy, neuropathy.
 XX
 PS Claim 1; SEQ ID NO 214; 701pp; English.
 XX
 CC The invention relates to novel genes and their fragments which are useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids and proteins
 CC are useful in the diagnosis, treatment and prevention of conditions
 CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,
 CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,
 CC infection, cataract, renal disorders, or endocrine disorders. The present
 CC sequence was used to illustrate the invention.
 XX
 SQ Sequence 1167 AA;
 Query Match 100.0%; Score 114; DB 6; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 KLGFFAHKKIPPEEKREKLEQ 22
 |||||
 Db 1146 KLGFFAHKKIPPEEKREKLEQ 1167
 RESULT 9
 ADQ19290
 ID ADQ19290 standard; protein; 1167 AA.
 XX
 AC ADQ19290;
 XX
 XX 26-AUG-2004 (first entry)
 DT
 XX Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.
 DE
 XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
 KW
 XX Homo sapiens.
 OS
 XX WO2004048938-A2.
 XX
 PN 10-JUN-2004.
 PD
 XX 26-NOV-2003; 2003WO-US038193.
 PF
 XX 26-NOV-2002; 2002US-0429739P.
 XX
 XX (PROT-) PROTEIN DESIGN LABS INC.
 PA
 XX

```

PI Aziz N, Gineburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
XX Example 2; SEQ ID NO 2109; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 1167 AA;

Query Match 100.0%; Score 114; DB 8; Length 1167;
Best Local Similarity 100.0%; Pred. No. 3.1e-07;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFFFAHKKIPBEKREKLEQ 22
DB 1146 KLGFFFAHKKIPBEKREKLEQ 1167

RESULT 10
ADC97404
ID ADC97404 standard; protein; 344 AA.
XX
AC ADC97404;
XX
DT 01-JAN-2004 (first entry)
XX
DE E. faecium protein sequence SEQ ID 7031.
XX
KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
KW abdominal-pelvic infection.
XX
OS Enterococcus faecium.
XX
PN US6583275-B1.
XX
PD 24-JUN-2003.
XX
PF 30-JUN-1998; 98US-00107532.
XX
PR 02-JUL-1997; 97US-0051571P.
PR 14-MAY-1998; 98US-0085598P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
XX WPI; 2003-799836/75.
DR N-PSDB; ADC93750.
XX
XX New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
XX Example 1; SEQ ID NO 7031; 243pp; English.
PS

```

```

XX
CC The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids are useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -
CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium proteins.
XX
XX Sequence 344 AA;

Query Match 45.6%; Score 52; DB 7; Length 344;
Best Local Similarity 47.1%; Pred. No. 47;
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 KLGFFFAHKKIPBEKRE 17
DB 13 RMNFFHYKKIPBEREQ 29

RESULT 11
ABG26508
ID ABG26508 standard; protein; 121 AA.
XX
AC ABG26508;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #26499.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSBQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS90695.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 56867; 103pp; English.
PS
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC

```

CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful for treating disorders
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 121 AA;

Query Match 44.7%; Score 51; DB 4; Length 121;
 Best Local Similarity 55.6%; Pred. No. 23;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 5 FAHKKIPDEEKREKLEQ 22
 | : ||| ||| : || : | :
 Db 5 FLYKKIKEEEEEEEEE 22

RESULT 12
 ABG11269
 ID ABG11269 standard; protein; 147 AA.
 AC
 XX ABG11269;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #11260.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 FN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 PF
 XX 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS75456.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 XX Claim 20; SEQ ID NO 41628; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a

CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 147 AA;

Query Match 44.7%; Score 51; DB 4; Length 147;
 Best Local Similarity 55.6%; Pred. No. 28;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 5 FAHKKIPDEEKREKLEQ 22
 | : ||| ||| : || : | :
 Db 5 FLYKKIKEEEEEEEEE 22

RESULT 13
 ABG07742
 ID ABG07742 standard; protein; 502 AA.
 AC
 XX ABG07742;
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #7733.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 FN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 PF
 XX 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS71929.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 XX Claim 20; SEQ ID NO 38101; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 502 AA;

Query Match 44.7%; Score 51; DB 4; Length 502;
 Best Local Similarity 55.6%; Pred. No. 95;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 5 FAHKKIPPEEKREKLEQ 22
 Db 5 FLYKKKEEEEEEE 22

RESULT 14
 AAU35717
 ID AAU35717 standard; protein; 856 AA.

XX AAU35717;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Helicobacter pylori cellular proliferation protein #30.
 XX
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.
 XX
 OS Helicobacter pylori.
 XX
 PN WO200170955-A2.
 XX

PD 27-SEP-2001.
 XX
 XX 21-MAR-2001; 2001WO-US009180.
 XX
 PR 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207727P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PI
 XX WPI; 2001-611495/70.
 DR
 DR N-PSDB; AAS53576.
 XX

XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 PT

XX Example 3; SEQ ID NO 11310; 511bp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes,
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also

CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 856 AA;

Query Match 44.7%; Score 51; DB 4; Length 856;
 Best Local Similarity 45.0%; Pred. No. 1.6e+02;
 Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
 QY 3 GFFAHHKIPPEEKREKLEQ 22
 Db 496 GEIYSKIPENKKKEELQR 515

RESULT 15
 ABU30742
 ID ABU30742 standard; protein; 856 AA.

XX ABU30742;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #16269.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KW Helicobacter pylori.
 OS
 XX WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX

PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 23-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 PI
 XX WPI; 2003-029926/02.
 DR
 DR N-PSDB; ACA34612.
 XX

XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 58666; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 856 AA;

Query Match 44.7%; Score 51; DB 6; Length 856;
 Best Local Similarity 45.0%; Pred. No. 1.6e+02;
 Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 GFFAAKKIPPEEKKEKLEQ 22

Db 496 GEIYSKIPENKKKEELQR 515

Search completed: April 6, 2005, 12:22:37
 Job time : 4.7823 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:17:21 ; Search time 2.74785 Seconds
(without alignments)
2658.060 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFFAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1418010 seqs, 33197259 residues

Total number of hits satisfying chosen parameters: 1418010

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	1167	US-10-741-601-531	Sequence 531, App
2	114	100.0	1177	US-10-741-601-532	Sequence 532, App
3	52.5	46.1	284	US-10-437-963-148869	Sequence 148869, App
4	52.5	46.1	1170	US-10-437-963-189133	Sequence 189133, App
5	52	45.6	137	US-10-424-599-277893	Sequence 277893, App
6	52	45.6	577	US-10-424-599-277888	Sequence 277888, App
7	51	44.7	856	US-09-812-350-2	Sequence 2, Appli
8	51	44.7	856	US-09-815-242-11310	Sequence 11310, A
9	51	44.7	856	US-10-282-122A-58666	Sequence 58666, A
10	50	43.9	1048	US-09-920-267C-9	Sequence 9, Appli
11	50	43.9	1048	US-10-211-462-189	Sequence 189, App
12	50	43.9	1048	US-10-720-323-9	Sequence 9, Appli
13	50	43.9	1048	US-10-482-029-20	Sequence 20, Appli

14	49.5	43.4	1388	16	US-10-408-765A-1139	Sequence 1139, Ap
15	49	43.0	107	16	US-10-437-963-204563	Sequence 204563,
16	49	43.0	108	15	US-10-424-599-155879	Sequence 155879,
17	49	43.0	413	14	US-10-083-357-1275	Sequence 1275, Ap
18	49	43.0	841	16	US-10-437-963-187432	Sequence 187432,
19	49	43.0	856	9	US-09-815-242-11489	Sequence 11489, A
20	49	43.0	856	15	US-10-335-977-7660	Sequence 7660, Ap
21	48	42.1	118	16	US-10-437-963-165708	Sequence 165708,
22	48	42.1	288	15	US-10-424-599-284529	Sequence 284529,
23	48	42.1	565	15	US-10-424-599-275593	Sequence 275593,
24	48	42.1	573	14	US-10-205-194-55	Sequence 55, Appl
25	47.5	41.7	126	16	US-10-437-963-115949	Sequence 115949,
26	47	41.2	49	15	US-10-424-599-184147	Sequence 184147,
27	47	41.2	122	15	US-10-335-977-8087	Sequence 8087, Ap
28	47	41.2	572	15	US-10-425-114-57970	Sequence 57970, A
29	47	41.2	646	15	US-10-425-114-70985	Sequence 70985, A
30	47	41.2	860	15	US-10-389-566-1868	Sequence 1868, Ap
31	47	41.2	885	16	US-10-437-963-160466	Sequence 160466,
32	46	40.4	118	16	US-10-767-701-50727	Sequence 50727, A
33	46	40.4	195	15	US-10-424-599-175252	Sequence 175252,
34	46	40.4	221	16	US-10-437-963-175346	Sequence 175346,
35	46	40.4	274	15	US-10-424-599-159588	Sequence 159588,
36	46	40.4	308	15	US-10-282-122A-73679	Sequence 73679, A
37	46	40.4	308	17	US-10-472-928-3922	Sequence 3922, Ap
38	46	40.4	363	16	US-10-437-963-182343	Sequence 182343,
39	46	40.4	479	9	US-09-971-309-64	Sequence 64, Appl
40	46	40.4	479	16	US-10-828-924-63	Sequence 63, Appl
41	46	40.4	753	14	US-10-032-585-7608	Sequence 7608, Ap
42	46	40.4	764	15	US-10-425-114-54331	Sequence 54331, A
43	46	40.4	1499	16	US-10-437-963-112408	Sequence 112408,
44	46	40.4	1603	16	US-10-828-924-62	Sequence 62, Appl
45	45	39.5	109	15	US-10-001-192A-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-10-741-601-531
; Sequence 531, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 531
; LENGTH: 1167
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-531

Query Match 100.0%; Score 114; DB 16; Length 1167;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGFFAHKKIPEEKREKLEQ 22
Db 1146 KLGFFAHKKIPEEKREKLEQ 1167

RESULT 2
US-10-741-601-532
; Sequence 532, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH STENOSIS, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741.601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-532

Query Match 100.0%; Score 114; DB 16; Length 1177;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPEEKREKLEQ 22
Db 1146 KLGFFAHKKIPEEKREKLEQ 1167

RESULT 3

US-10-437-963-148869
; Sequence 148869, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437.963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 148869
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Oryza sativa

; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_49258C.1.pep

US-10-437-963-148869

Query Match 46.1%; Score 52.5; DB 16; Length 284;
Best Local Similarity 57.9%; Pred. No. 45;
Matches 11; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 LGFFAHKKIPEEKREKLEQ 20
Db 131 LGLF-DRRIPEEEEEKEM 148

RESULT 4

US-10-437-963-189133
; Sequence 189133, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437.963

; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 189133
; LENGTH: 1170
; TYPE: PRT
; ORGANISM: Oryza sativa

; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85670C.1.pep

US-10-437-963-189133

Query Match 46.1%; Score 52.5; DB 16; Length 1170;
Best Local Similarity 57.9%; Pred. No. 1.9e+02;
Matches 11; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 LGFFAHKKIPEEKREKLEQ 20
Db 1017 LGLF-DRRIPEEEEEKEM 1034

RESULT 5

US-10-424-599-277893
; Sequence 277893, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 277893
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Glycine max

; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_92960C.1.pep

US-10-424-599-277893

Query Match 45.6%; Score 52; DB 15; Length 137;
Best Local Similarity 41.2%; Pred. No. 25;
Matches 7; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREKLEQ 22
Db 93 SHRRVPEKSEEEELSE 109

RESULT 6

US-10-424-599-277888
; Sequence 277888, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 277888
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Glycine max

; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(577)

```
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_92956C.1.pep
US-10-424-599-277888

Query Match      45.6%; Score 52; DB 15; Length 577;
Best Local Similarity 41.2%; Pred. No. 1.le+02;
Matches 7; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

Qy 6 AHKKIPEEKREKLEQ 22
:|::|||: :|||:
Db 93 SHRVPEKSEEBLEE 109

RESULT 7
US-09-812-350-2
; Sequence 2, Application US/09812350
; Publication No. US20020053097A1
; GENERAL INFORMATION:
; APPLICANT: Lindquist, Susan
; APPLICANT: Queitsch, Christine
; APPLICANT: Vierling, Elizabeth
; TITLE OF INVENTION: Transgenic Plants Containing Heat Shock Protein
; FILE REFERENCE: P01979US2
; CURRENT APPLICATION NUMBER: US/09/812,350
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/190,769
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: US 60/198,116
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-812-350-2

Query Match      44.7%; Score 51; DB 9; Length 856;
Best Local Similarity 45.0%; Pred. No. 2.le+02;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHKKIPEEKREKLEQ 22
| :||| :|::||:
Db 496 GEIYSKIPENKKKEBELQR 515

RESULT 8
US-09-815-242-11310
; Sequence 11310, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58666
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-282-122A-58666
```

```
; OTHER INFORMATION:
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11310
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11310

Query Match      44.7%; Score 51; DB 9; Length 856;
Best Local Similarity 45.0%; Pred. No. 2.le+02;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHKKIPEEKREKLEQ 22
| :||| :|::||:
Db 496 GEIYSKIPENKKKEBELQR 515

RESULT 9
US-10-282-122A-58666
; Sequence 58666, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58666
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-282-122A-58666
```

```

RESULT 11
US-10-211-462-189
  / Sequence 189, Application US/10211462
  / Publication No. US20040033495A1
  / GENERAL INFORMATION:
  / APPLICANT: Murray, Richard
  / APPLICANT: Glynnne, Richard
  / APPLICANT: Watson, Susan R.
  / APPLICANT: Aziz, Nataasha
  / APPLICANT: Eos Biotechnology, Inc.
  / TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
  / TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
  / FILE REFERENCE: 018501-006200US
  / CURRENT APPLICATION NUMBER: US/10/211,462
  / CURRENT FILING DATE: 2003-02-13
  / PRIOR APPLICATION NUMBER: US 09/784,356
  / PRIOR FILING DATE: 2001-02-14
  / PRIOR APPLICATION NUMBER: US 09/791,390
  / PRIOR FILING DATE: 2001-02-22
  / PRIOR APPLICATION NUMBER: US 60/310,025
  / PRIOR FILING DATE: 2001-08-03
  / PRIOR APPLICATION NUMBER: US 60/334,244
  / PRIOR FILING DATE: 2001-11-29
  / NUMBER OF SEQ ID NOS: 230
  / SOFTWARE: PatentIn Ver. 2.1
  / SEQ ID NO 189
  / LENGTH: 1048
  / TYPE: PRT
  / ORGANISM: Homo sapiens
US-10-211-462-189

```

RESULT 14
US-10-408-765A-1139
; Sequence 1139, Application US/10408765A
; Publication No. US20040101874A1

; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Martock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 66088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1139
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1139

Query Match 43.4%; Score 49.5; DB 16; Length 1388;
Best Local Similarity 57.9%; Pred. No. 5.5e+02;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

Qy 4 FFAH---KKIPEEKREEK 19
Db 366 FFAHLLQKVLAEERKQK 384

RESULT 15

US-10-437-963-204563
; Sequence 204563, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 204563
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_99638C.1.pep
US-10-437-963-204563

Query Match 43.0%; Score 49; DB 16; Length 107;
Best Local Similarity 47.4%; Pred. No. 49;
Matches 9; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPEEKREEK 19
Db 82 KLGFSPKPKKKKKKKKK 100

Search completed: April 6, 2005, 12:53:10
Job time : 3.74785 secs

Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 0.938919 Seconds
(without alignments)
1749.117 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114

Sequence: 1 KLGFPAHKTIPEEKREKLEQ 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:**

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	45.6	344	4 US-09-107-532A-7031	Sequence 7031, Ap
2	50	43.9	32	1 US-08-214-770-2	Sequence 2, Appli
3	50	43.9	32	5 PCT-US95-02885-2	Sequence 2, Appli
4	50	43.9	1048	4 US-09-949-016-6192	Sequence 6192, Ap
5	50	43.9	1149	4 US-09-949-016-7401	Sequence 7401, Ap
6	48	42.1	573	2 US-08-745-934-4	Sequence 4, Appli
7	48	42.1	573	3 US-09-147-009-7	Sequence 7, Appli
8	47.5	41.7	2343	3 US-09-324-867-2	Sequence 2, Appli
9	46	40.4	308	4 US-09-583-110-2953	Sequence 2953, Ap
10	46	40.4	331	4 US-09-107-433-2823	Sequence 2823, Ap
11	46	40.4	450	4 US-09-538-092-493	Sequence 493, Ap
12	46	40.4	479	3 US-09-446-504-64	Sequence 64, Appl
13	46	40.4	479	3 US-07-112-266-64	Sequence 64, Appl
14	46	40.4	755	4 US-09-248-796A-14733	Sequence 14733, A
15	45	39.5	109	3 US-09-091-725-31	Sequence 31, Appl
16	45	39.5	476	4 US-09-248-796A-15008	Sequence 15008, A
17	44	38.6	46	1 US-08-200-016-3	Sequence 3, Appli
18	44	38.6	93	1 US-07-987-272A-7	Sequence 7, Appli
19	44	38.6	93	1 US-07-987-272A-16	Sequence 16, Appl
20	44	38.6	93	1 US-08-385-241-1	Sequence 1, Appli
21	44	38.6	93	4 US-09-214-272-2	Sequence 2, Appli
22	44	38.6	93	4 US-09-806-382A-3	Sequence 3, Appli
23	44	38.6	200	4 US-09-702-705-789	Sequence 789, App
24	44	38.6	200	4 US-09-736-457-789	Sequence 789, App
25	44	38.6	200	4 US-09-614-124B-789	Sequence 789, App
26	44	38.6	200	4 US-09-671-325-789	Sequence 789, App
27	44	38.6	200	4 US-09-589-184-789	Sequence 789, App

28	44	38.6	200	4 US-09-658-824-789	Sequence 789, App
29	44	38.6	207	4 US-09-702-705-1667	Sequence 1667, Ap
30	44	38.6	207	4 US-09-736-457-1667	Sequence 1667, Ap
31	44	38.6	207	4 US-09-614-124B-1667	Sequence 1667, Ap
32	44	38.6	207	4 US-09-671-325-1667	Sequence 1667, Ap
33	44	38.6	207	4 US-09-658-824-1667	Sequence 1667, Ap
34	44	38.6	228	4 US-09-949-016-10496	Sequence 10496, A
35	44	38.6	481	4 US-09-248-796A-17118	Sequence 17118, A
36	44	38.6	745	4 US-09-248-796A-17272	Sequence 17272, A
37	44	38.6	995	4 US-09-362-842-2	Sequence 2, Appli
38	43.5	38.2	765	1 US-08-425-061-19	Sequence 19, Appl
39	43.5	38.2	765	2 US-08-825-886-19	Sequence 19, Appl
40	43.5	38.2	765	4 US-08-989-890-19	Sequence 19, Appl
41	43.5	38.2	900	1 US-08-425-061-20	Sequence 20, Appl
42	43.5	38.2	900	2 US-08-825-886-20	Sequence 20, Appl
43	43.5	38.2	900	4 US-08-989-890-20	Sequence 20, Appl
44	43.5	38.2	914	1 US-08-425-061-21	Sequence 21, Appl
45	43.5	38.2	914	2 US-08-825-886-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-107-532A-7031
; Sequence 7031, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107.532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 7031:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...344
; SEQUENCE DESCRIPTION: SEQ ID NO: 7031:
US-09-107-532A-7031

Qy 1 KLGFPAHKKIPPEEKREKLE 21
:|||||:|:|:|:|:|:
Db 1017 RMGFFKVRPPQEQEREQLQ 1037

RESULT 5

US-09-949-016-7401
; Sequence 7401, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7401
; LENGTH: 1149
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7401

Query Match 43.9%; Score 50; DB 4; Length 1149;
Best Local Similarity 38.1%; Pred. No. 84;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
Qy 1 KLGFPAHKKIPPEEKREKLE 21
:|||||:|:|:|:|:|:
Db 1118 RMGFFKVRPPQEQEREQLQ 1138

RESULT 6

US-08-745-934-4
; Sequence 4, Application US/08745934
; Patent No. 5861496
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: HUMAN SQUALENE EPOXIDASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/745,934
; FILING DATE: Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0151 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555

; TELEFAX: (415) 845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1083804
; US-08-745-934-4

Query Match 42.1%; Score 48; DB 2; Length 573;
Best Local Similarity 52.9%; Pred. No. 80;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Qy 2 LGFFAHHKIPPEEKREE 18
:|||||:|:|:|:|:|:
Db 75 IGFFWAKSPPESEKKEQ 91

RESULT 7

US-09-147-009-7
; Sequence 7, Application US/09147009
; Patent No. 6153815
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for Raising Squalene Levels in Plants
; TITLE OF INVENTION: and DNA Sequences Used Therefor
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/147,009
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Rattus norvegicus
; TISSUE TYPE: kidney
; CELL LINE: NRK
; IMMEDIATE SOURCE:
; LIBRARY: pCD2 library of H. Okayama
; CLONE: Tb-1
; PUBLICATION INFORMATION:
; AUTHORS: Sakakibara, J.
; AUTHORS: Watanabe, R.
; AUTHORS: Kanai, R.
; AUTHORS: Ono, T.
; TITLE: Molecular cloning and expression of rat
; TITLE: squalene epoxidase
; JOURNAL: J. Biol. Chem.
; VOLUME: 270
; ISSUE: 1
; PAGES: 17-20
; DATE: 1995
; US-09-147-009-7

Query Match 42.1%; Score 48; DB 3; Length 573;
Best Local Similarity 52.9%; Pred. No. 80;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LGFAHKKIPBEKREE 18
:|||||:|||||:
Db 75 IGFWAKSPSEKKEQ 91

RESULT 8

US-09-324-867-2
; Sequence 2, Application US/09324867A
; Patent No. 6251632
; GENERAL INFORMATION:
; APPLICANT: Lillicrap, David
; APPLICANT: Cameron, Cherie
; APPLICANT: No. 6251632ley, Colleen
; APPLICANT: Horrocks, L. Suzanne Hoyle
; APPLICANT: Hough, Christine
; TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
; FILE REFERENCE: 1669-0010002/JAG/BJD
; CURRENT APPLICATION NUMBER: US/09/324, 867A
; CURRENT FILING DATE: 1999-06-03
; EARLIER FILING DATE: 1998-03-141
; EARLIER FILING DATE: 1998-03-059
; EARLIER FILING DATE: 1997-03-953
; EARLIER FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2343
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-324-867-2

Query Match 41.7%; Score 47.5; DB 3; Length 2343;
Best Local Similarity 40.6%; Pred. No. 3.7e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 13; Gaps 1;

Qy 4 FFAH-----KKIPBEKREKLEQ 22
:|||||:|||||:
Db 1183 FFLANLVQNDVTYQKSPETIERKELTQ 1214

RESULT 9

US-09-583-110-2953
; Sequence 2953, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 2953
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-2953

Query Match 40.4%; Score 46; DB 4; Length 308;
Best Local Similarity 36.8%; Pred. No. 82;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHKKIPBEKREKLE 21
:|||||:|||||:
Db 113 GLYNHRLFKDEERKEKVKQ 131

RESULT 10

US-09-107-433-2623
; Sequence 2623, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
; COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2623:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
; FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...331
; SEQUENCE DESCRIPTION: SEQ ID NO: 2623:
US-09-107-433-2623

Query Match 40.4%; Score 46; DB 4; Length 331;
Best Local Similarity 36.8%; Pred. No. 88;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHKKIPBEKREKLE 21
:|||||:|||||:
Db 136 GLYNHRLFKDEERKEKVKQ 154

RESULT 11

US-09-538-092-493
; Sequence 493, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352

US-09-091-725-31
 ; Sequence 31, Application US/09091725
 ; Patent No. 6329141
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT:
 ; TITLE OF INVENTION: Improved methods for transforming Phaffia
 ; TITLE OF INVENTION: and recombinant DNA for use therein
 ;
 ; NUMBER OF SEQUENCES: 51
 ;
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morrison & Foerster llp
 ;

```

; STREET: 2000 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: United States of America
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,725
; FILING DATE: 23-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95203620.0
; FILING DATE: 22-DEC-1995
; APPLICATION NUMBER: EP 96200943.7
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: E. Victor Donahue
; REGISTRATION NUMBER: 35,492
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-091-725-31

```

```

Query Match      39.5%; Score 45; DB 3; Length 109;
Best Local Similarity 52.6%; Pred. No. 40;
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

```

```

Qy      3 GPPAHKKIPEEKKEKLE 21
Db      76 GASADASAPAEKKEKAE 94

```

```

Search completed: April 6, 2005, 12:24:47
Job time : 1.93892 secs

```